Deliverable D7.3

Data Management Plan–v2

DRAFT

Editor(s):	Cristina Martínez, Juncal Alonso (TECNALIA)
Responsible Partner:	TECNALIA Research and Innovation
Status-Version:	Final – v1.0
Date:	30.10.2026
Type:	DMP
Distribution level (SEN, PU):	PU

Project Number:	101120688
Project Title:	EMERALD

Title of Deliverable:	D7.3 Data Management Plan–v2
Due Date of Delivery to the EC	31.10.2026

Work package responsible for the Deliverable:	WP7 – Project management
Editor(s):	Cristina Martínez, Juncal Alonso (TECNALIA)
Contributor(s):	IONOS, CF, FABAR, ONS, CXB
Reviewer(s):	Björn Fanta (FABAR)
Approved by:	All Partners
Recommended/mandatory readers:	All WPs

Abstract:	Final version of the data management strategy of the project, and the established procedures and criteria to guarantee ethics and privacy in all project activities.
Keyword List:	Data Management Plan, OpenAIRE, Repository, Metadata, FAIR data
Licensing information:	This work is licensed under Creative Commons Attribution-ShareAlike 4.0 International (**CC BY-SA 4.0 DEED** https://creativecommons.org/licenses/by-sa/4.0/)
Disclaimer	Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them.

DRAFT

Document Description

Version	Date	Modifications Introduced	
		Modification Reason	Modified by
v0.1	02.10.2026	First draft version	Cristina Martínez, Juncal Alonso (TECNALIA)
v0.2	08.10.2026	Comments and suggestions received by consortium partners	IONOS, CF, FABAR, ONS, CXBTECH
v0.3	28.10.2026	QA review	Björn Fanta (FABAR)
v0.4	29.10.2026	Address the comments from the QA review	Cristina Martínez, Juncal Alonso (TECNALIA)
v1.0	30.10.2026	Submitted to the European Commission	Cristina Martínez (TECNALIA)

DRAFT

(cc) BY-SA

Table of contents

Terms and abbreviations 6
Executive Summary 7
1 Introduction 8
1.1 About this deliverable 8
1.2 Document structure 8
1.3 Overview of updates compared to D7.2 9
2 Data Summary 10
2.1 Overview of EMERALD's key results and implementation environments 10
2.2 Data Classification 12
2.3 Interviews and focus groups data 12
2.4 Pilot data 13
2.4.1 Pilot 1: Certification of public Cloud Services - IONOS 14
2.4.2 Pilot 2: Certification of public Cloud Services – CloudFerro 14
2.4.3 Pilot 3: Certification of public Cloud Services - Fabasoft 14
2.4.4 Pilot 4: Certification of hybrid Cloud-edge environments for the financial sector 14
2.5 Scientific publications 15
2.6 Project deliverables 15
2.7 Other publications 18
2.8 Data used during the CI/CD process 19
2.9 Data stored in the Repository of metrics 19
3 FAIR Data 20
3.1 Making data findable, including provisions for metadata 20
3.2 Making data openly accessible 21
3.2.1 Repositories 21
3.2.2 Metadata 23
3.2.3 Data related to the pilots 24
3.3 Making data interoperable 26
3.4 Increase data re-use 26
4 Other Research Outputs 27
5 Allocation of Resources 28
6 Data Security 29
6.1 Personal data protection management 29
6.2 Security management 30
7 Ethics 32
8 Conclusions 33

DRAFT

CC BY-SA

9 Referencias 34
APPENDIX A: General Informed Consent Form 36
APPENDIX B: Information for Conducting Interviews in WP4 38

List of tables

TABLE 1. OVERVIEW OF UPDATES WITH RESPECT TO D7.2 [1] 9
TABLE 2. PUBLIC DELIVERABLES IN EMERALD 16

List of figures

FIGURE 1. EMERALD INTEGRATION AND PRODUCTION ENVIRONMENTS [6] 12
FIGURE 2. EMERALD PILOTS 13
FIGURE 3. EMERALD PRIVATE DOCUMENT REPOSITORY IN FABASOFT CLOUD 21
FIGURE 4. PUBLIC GITLAB REPOSITORY FOR EMERALD 22
FIGURE 5. ZENODO COMMUNITY FOR EMERALD 22
FIGURE 6. EMERALD METRICS IN THE COLLABORATIVE METRIC REPOSITORY 23

DRAFT

CC BY-SA

Terms and abbreviations

AI	Artificial Intelligence
AI-SEC	AI Security Evidence Collector
AMOE	Assessment and Management of Organizational Evidence
API	Application Programming Interface
CI/CD	Continuous integration/Continuous delivery
CaaS	Certification as a Service
CERN	European Organization for Nuclear Research
CSP	Cloud Service Provider
DoA	Description of Action
DMP	Data Management Plan
DPO	Data Protection Officer
EC	European Commission
ETSI	European Telecommunications Standards Institute
EUPL	European Union Public Licence
FAIR	Findable, Accessible, Interoperable and Reusable
GA	Grant Agreement
GDPR	General Data Protection Regulation
IaaS	Infrastructure as a Service
IP	Intellectual Property
IPR	Intellectual Property Rights
JSON	JavaScript Object Notation
KPI	Key Performance Indicators
KR	Key Result
MARI	Mapping Assistant for Regulations with Intelligence
ML	Machine Learning
NLP	Natural Language Processing
OPA	Open Policy Agent
OSS	Open Source Software
PA	Public Administrations
PaaS	Platform as a Service
PC	Project Coordinator
PoC	Proof of Content
RBAC	Role-Based Access Control
RCM	Repository of Controls and Metrics
SaaS	Software as a Service
SAB	Security Advisory Board
SME	Small and Medium Enterprise
SUS	System Usability Scale
SW	Software
TM	Technical Manager
TRL	Technology Readiness Level
TWS	Trustworthiness System
UI/IX	User Interface/User Experience
YAML	YAML Ain’t Markup Language
WP	Work Package

DRAFT

(cc) BY-SA

Executive Summary

As part of EMERALD project management, Task 7.4 “Data Management” is in charge of defining and revising the Data Management Plan (DMP) of the project.

The present report is the final version of the DMP which describes the data used and the data produced by the project along with the procedures designed for data life-cycle management. The report builds upon deliverable D7.2 [1] (M6) reflecting updates mainly in the final data types generated in the project.

This DMP sets out the methodology followed by the EMERALD consortium members throughout the project to handle the data used and created in the project, including its collection, generation, storage and preservation. Throughout the duration of the project, the initial version of the DMP has undergone regular revisions to adapt it to the project’s needs and to changes in the data used and generated.

The report presents the categories of the data created and processed (including storage) in EMERALD: interview and focus group data, pilot data, scientific publications, public deliverables, other publications, data used during CI/CD processes and data stored in the repository of metrics.

The document also presents, following the EC template [2], how these different types of data have been collected, who the main beneficiaries are, and how they are stored and managed to follow the FAIR principles (Findable, Accessible, Interoperable and Reusable).

The plan also identifies the roles and responsibilities of key partners and participants in the project and describes the data security and privacy measures adopted in the project to ensure that security sensitive information is not disclosed in the project data. Finally, ethical aspects that should be taken into consideration in the context of EMERALD are also described as part of the plan.

DRAFT

CC BY-SA

1 Introduction

1.1 About this deliverable

This deliverable is the final result of task T7.4 "Data Management" and details the final Data Management Plan (DMP) of the EMERALD project. This plan defines the procedures and criteria established in the project in relation to the data management strategy, including the ways in which data are collected, generated and processed throughout the life of the project, while ensuring ethics and privacy in project activities and results.

The EMERALD project handles different types of data: interview and focus group data, pilot data, scientific publications, public deliverables, other publications, data used during CI/CD processes and data stored in the repository of metrics. The document presents how these different types of data are collected and stored, and how EMERALD manages them to make them findable, accessible, interoperable and reusable.

The plan follows the structure set out in the Horizon Europe template for a Data Management Plan (DMP) [2]. First, the plan identifies the types of data that will be used in the project. Second, it outlines how the data of the project will comply with the FAIR data principles. Third, it discusses about the management of other research outputs that are generated/re-used in the project. Fourth, the roles and responsibilities of the partners and key participants responsible for data management and quality assurance are identified. Fifth, it describes the data security and data privacy measures adopted in the project to prevent disclosure of privacy-sensitive or security-sensitive information. Finally, ethics aspects in relation to the data are described.

The initial version of the DMP produced at M6 has been periodically reviewed by the project coordinator and updated by all partners as necessary as the project progressed. The DMP was also regularly checked during the project's interim technical reports, which were produced every 6 months. In order to enable continuous monitoring and facilitate partners' contributions to the DMP, the interim version of the DMP was accessible in the project Fabasoft repository so as all the partners could propose updates and suggestions for the document.

The present report is the final version of the plan which describes the data used and the data produced by the project along with the procedures designed for data life-cycle management. The report builds upon deliverable D7.2 [1] (M6) reflecting updates mainly in the final data types generated in the project.

1.2 Document structure

The document is structured as follows.

Section 2 presents a summary of what the purpose of the data collection and generation is in the case of EMERALD. Section 3 explains how the data and metadata have been made fair, and thus findable, accessible, interoperable and reusable. Section 4 presents the management of other research outputs that were generated/re-used in the project. Section 5 briefly explains how the resources were allocated. Sections 6 and 7 focus on the security and ethical aspects, respectively. Finally, section 8 concludes the document providing a summary of the report.

APPENDIX A: General Informed Consent Form includes the form to be used to gain signed informed consent from data subjects involved in project activities to authorise the use of images and/or voice if needed.

APPENDIX B: Information for Conducting Interviews in WP4 includes the contents of the Participant Information Sheet, the Informed Consent, and Data Protection Information used in WP4 for conducting the interviews.

(cc) BY-SA

1.3 Overview of updates compared to D7.2

This section provides a summary of the updates and additions included in this document, the final version of the DMP, compared to its initial version (deliverable D7.2 [1]) in M6. Table 1 provides an overview of these changes.

Table 1. Overview of updates with respect to D7.2 [1]

Section	Changes
Executive Summary	Minor changes in the text.
1.Introduction	Minor changes in the text and addition of the subsection "1.3 Updates from D7.2".
2. Data Summary	Minor changes in the text. A new data type (data stored in the Repository of metrics) was added and documented in the new subsection 2.9.
3. FAIR Data	Minor changes in the text. Detailed information on the repositories and its associated metadata. Data related to the pilots reviewed and updated.
4. Other Research Outputs	Minor changes in the text.
6. Data Security	*Keycloak* was chosen as the open-source Identity and Access Management (IAM) solution.
7. Ethics	Minor changes in the text.
8. Conclusions	Conclusions updated.

DRAFT

CC BY-SA

2 Data Summary

The research activities of the EMERALD project require the management of various types of data, both those used in the project and those generated by it. In this section, we first provide an overview of EMERALD's key results and its implementation environments and then describe the types of data resulting from the project's activities, together with the description of their purpose. This data are open data that will be shared amongst EMERALD participants as well as with audiences outside the project.

2.1 Overview of EMERALD's key results and implementation environments

The aim of the EMERALD project is to transform the landscape of cloud-based services, focusing on developing a new framework to enhance security and efficiency for both large and SMEs. With a commitment to developing an agile certification process, EMERALD supports Cloud Service Providers, Customers and Auditors in the certification process, fostering the adoption of cloud services and ensuring they are accessible, secure, and beneficial for all stakeholders involved.

EMERALD leverages the findings of the H2020 MEDINA project [3], starting from TRL 5 and advancing in the EMERALD core to TRL 7. The focus is on evidence management components that are exploited for developing a continuous certification approach. In particular, the following components from MEDINA have evolved [4]:

- Catalogue of Controls and Metrics, which evolved to the *Repository of Controls and Metrics (RCM)*
- Metric Assistant, which evolved to the *Mapping Assistant for Regulations with Intelligence* (*MARI*)
- *Trustworthiness system*
- *Orchestrator*
- *AMOE*
- *Codyze*
- *Clouditor Discovery*

In addition, two other components have been integrated:

- *eknows-e3*
- *AI-SEC*
- *EMERALD-UI*

To realize the main goal of the project, EMERALD developed the following KRs, as described in the DoA [5]:

- **KR1 EXTRACT. Evidence extraction from cloud service:** A framework to continuously extract knowledge on various layers of the cloud service (infrastructure, code, business processes) and prepare suitable evidence based on them. This result covers the improvements on existing evidence extraction tools and concepts of MEDINA, such as AMOE (Assessment and Management of Organisational Evidence). The framework works on utilizing information and verifies the implementation of technical and organisational measures. The tools enable different levels of abstraction, from low level such as source code to higher levels, such as policies and procedures.
- **KR2 CERTGRAPH: Certification graph:** A graph-based structure, the certification graph, to consolidate all necessary information of the service and make it easily query-able. The graph-based approach allows storing and linking heterogeneous information

(cc) BY-SA

extracted from different evidence sources. Furthermore, linking allows to create additional nodes in the graph that aggregate individual aspects and fragments of information to a higher-level of combined evidence, while maintaining traceability back to information sources.

- **KR3 OPTIMA. Optimized metric selection:** An intelligent system to select an optimized set of metrics that can be measured to demonstrate compliance to the selected certification scheme. One of such optimizations could be the maximum amount of re-used evidence.
- **KR4 MULTICERT. Cloud certifications:** A tool to assess chosen metrics based on information stored in the certification graph and to evaluate the final certificate decision.
- **KR5 AIPOC. Proof of Concept for AI-based certifications**: By transferring the innovation results to upcoming AI certification schemes, EMERALD establishes a Proof of Concept (PoC) on how to scale the Certification as a Service (CaaS) approach to cloud-based AI systems.
- **KR6 EMERALD UI/UX. User experience for complexity reduction:** A user interaction concept and conducted studies to show what information each user needs in an audit process. The concept shall lead to a user interface (UI), which is tailored to the users' needs during all stages of an audit and guides them through the process of identifying problems top down – from high level requirements down to specific implementation in documents (e.g., policies) or technical specifications.
- **KR7 INTEROP. Interoperable assessment, evidence and catalogue data:** EMERALD will provide an interoperability layer, including the trustworthy system, assessment results and catalogue data. Security schemes are prone to change and thus updates would be required. EMERALD aims to mitigate this by incorporating the scheme data in a standardized format such as OSCAL (Open Security Controls Assessment Language). Furthermore, EMERALD aims at providing interoperability at the trustworthy evidence layer by evaluating usage of the European Blockchain Services Infrastructure (EBSI) for its trustworthiness system.
- **KR8 PILOTS. Industrial pilots:** Involvement of realistic use cases by potential applicants of EMERALD. This is key to derive and validate the proposed contents of the project objectives. PILOTS is responsible for providing these real-world application examples and test data. The data will be forwarded to the evidence extraction stakeholders, so the components can be fine-tuned to improve quality of the results.
- **KR9 DECAS. Dissemination, exploitation, communication and standardization:** Dissemination and communication of the project results via multiple channels, relevant conferences (e.g., ETSI security conference) and the scientific community. Exploitation of the project achievements by the technical and pilot partners. Standardization activities to discuss, verify and deepen the project findings with standardization bodies. Concepts for continuous use and deployment after the project has finished need to be prepared and documented.

Out of all the Key Results envisioned for EMERALD, we have identified and analysed those that are most data-oriented, either by using, generating or analysing data, and none of them involve personal data:

- **KR1 EXTRACT**: The components of this KR do not extract any personal data from the CSP services.
- **KR3 OPTIMA**: The components of this KR do not involve the processing of any personal data.
- **KR5 AIPOC**: The PoC for AI based certification does not involve the processing of personal data.

CC BY-SA

- **KR6 EMERALD UI/UX**: Interviews and focus groups were organized to gather UI/UX requirements. The personal data of the people involved in them was anonymized.
- **KR8 PILOTS**: Category II pilot, focused on the financial sector, might seem especially relevant, but it should be noted that this pilot does not involve the processing of personal data.

The EMERALD CaaS framework is supported by a Kubernetes cluster through two different environments: **Integration** and **Production** (see Figure 1). The components are first deployed in the Integration environment in containerized form. Once the integration tests have been passed, components are promoted to the Production environment; one production environment has been set up for each of the pilots. All the environments are built in a seven-node ***Kubernetes***[1] cluster over an ***OpenStack***[2] platform. They are deployed at TECNALIA's DevOps infrastructure and include the different components developed by the EMERALD technological partners. Interested readers can find more detailed information in D1.9 [6].

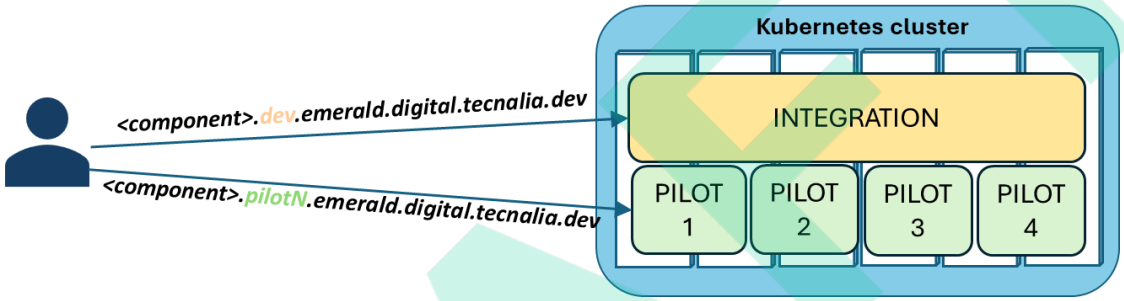

Figure 1. EMERALD integration and production environments [6]

2.2 Data Classification

The EMERALD project handles different types of data, which are described in the following sections:

- Data from interviews and focus groups
- Data related to the pilots
- Data related to scientific publications
- Project deliverables
- Other publications
- Data used during the CI/CD process
- Data stored in the Repository of metrics

2.3 Interviews and focus groups data

Several interviews were conducted with pilot owners within WP4 – User Interaction and User Experience development of the EMERALD Project. The goal of these interviews was to elicit requirements from EMERALD target groups, such as auditors/chief information security managers/compliance managers, etc., necessary for developing the integrated EMERALD UI.

The main goal was to elicit in-depth insights about the work of [auditors/chief information security managers/compliance managers] in relation to continuous cloud auditing processes. People involved were those who know how the preparation processes for an audit look like at

[1] https://kubernetes.io/docs/home/
[2] https://www.openstack.org

(cc) BY-SA

their organisation, which documents to prepare or how to implement new controls to satisfy the standard to be audited. Between 3-5 interviews were conducted per pilot partner.

The interviewees received a participant information sheet and were asked to sign an informed consent, as explained in section 6.1. The interviews were recorded, and information shared with the project members (only answers without any personal information). The description of the process and the analysis of the data was published in D4.1 [7], D4.2 [8] and in other scientific publications.

During the last months of the project several activities were also conducted to validate the EMERALD CaaS framework in the context of WP5. The UI/UX validation was conducted in structured sessions, each comprising three sequential steps: a guided series of user tasks performed directly in the EMERALD interface, completion of the SUS questionnaire, and a concluding semi-structured interview. The SUS questionnaire was administered via an online survey tool called *LimeSurvey*[3], and responses were collected directly through the platform. Interview notes were taken in written form by the pilot owner during each session.

No personal data was collected as part of the validation. In accordance with the guidance provided to pilot owners, results were to be reported without reference to individual participants by name. All documentation was submitted for the purpose of preparing the Validation Report (D5.6 [9]).

2.4 Pilot data

The aim of this section is to establish what the purpose of data collection and generation is within the EMERALD pilots, what the sources of this data are, whether they are relevant, and how this is important for realising the EMERALD Key Results.

To validate the EMERALD framework, two main categories of pilots were identified (see Figure 2). The goal of these pilots is to illustrate how EMERALD will enable regulated industries to move to cloud-first/native environments while deriving individual compliance level as per industry or customer requirement and provide services fulfilling business needs. The Pilots are a key pillar of EMERALD as they validate (in WP5) the work done in the technical work packages (WP1-WP4).

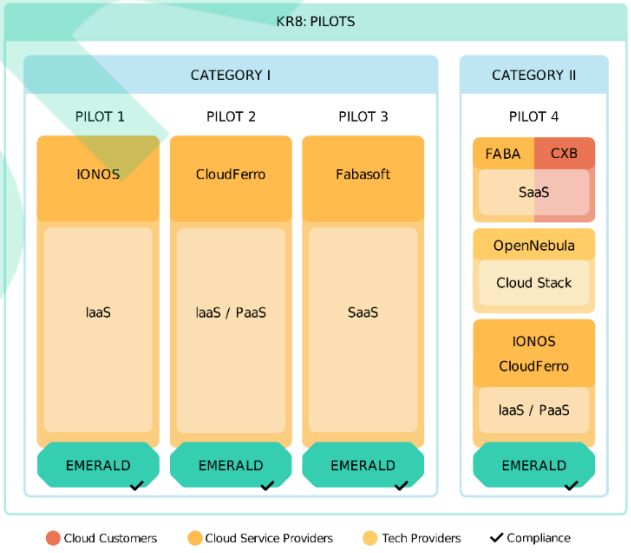

Figure 2. EMERALD Pilots

[3] https://www.limesurvey.org

CC BY-SA

The first category of pilots includes three use cases, which aim at demonstrating Certification as a Service with EMERALD, on the level of cloud services for IaaS (Pilot 1, IONOS), IaaS/PaaS (Pilot 2, CloudFerro) and SaaS (Pilot 3, Fabasoft). All three describe the path for integrating EMERALD tools from for European CSPs and set their focus to public cloud environments.

The second category of pilots aims at the certification of hybrid cloud-edge environments for the financial sector. The main driver of this category definition is the partner CaixaBank (Pilot 4), which holds a large number of on-premises services and is trying to expand this into the field of public clouds, i.e., using SaaS or IaaS providers. The application of EMERALD will ensure the real-time assessment of several cloud services, validating that they are compliant with the controls defined in a specific security framework.

The following subsections include a description of the data managed by each of the Pilots at the final stage of the project.

2.4.1 Pilot 1: Certification of public Cloud Services - IONOS

This pilot does not involve the processing of personal data. The data managed by this pilot includes:

- Policy documents for the *AMOE* tool: Security policies, protocols, and procedures.
- Sample datasets and test cases for validating data handling processes.
- Simulated compliance scenarios for testing the system's performance.

2.4.2 Pilot 2: Certification of public Cloud Services – CloudFerro

This pilot does not involve the processing of personal data. The data managed by this pilot includes:

- Data about test environments for IaaS and PaaS (dedicated OpenStack projects on CloudFerro's public cloud) which is collected by *Clouditor* via OpenStack API e.g. available virtual machines and their configuration.
- Anonymized versions of policies, procedures, certificates, and other documents used by *AMOE* for evidence collection .

2.4.3 Pilot 3: Certification of public Cloud Services - Fabasoft

This pilot does not involve the processing of personal data. The data managed by this pilot includes:

- app.telemetry logs: app.telemetry collects definable log data from various sources. This information is then checked against predefined values in order to notify operators in case of errors. These logs are used as evidence in EMERALD.
- (OPTIONAL) Objects in Fabasoft Cloud/Fabasoft Folio: Data is generally stored as an object in Fabasoft Cloud. They have predefined object properties that can be collected and checked. For EMERALD, objects and their properties shall be used as evidence.
- Synthetic policy documents for the *AMOE* tool.

2.4.4 Pilot 4: Certification of hybrid Cloud-edge environments for the financial sector

This pilot does not involve the processing of personal data. In case some minor accidental processing of personal data took place, it would be handled in accordance with applicable law. However, it is possible that data from pilots 1, 2 and 3 will be forwarded to pilot 4, as the findings

CC BY-SA

of pilots 1-3 are – on occasion – of interest to parts of pilot 4. In that case, pilot 4 will have to address these data as pilots 1-3 did.

The data managed by this pilot includes:

- Policy documents: about 20 synthetic PDF policies and procedures of a simulated CaixaBank supplier, assessed by the *AMOE* tool.
- Resource configuration data: virtual machines, storage, network interfaces and user identities of an OpenNebula test environment, collected by the OpenNebula evidence collector.
- Evidence and assessment results stored in the EMERALD Evidence Store.
- Audit records: audit scope, auditor decisions and comments, and the audit report.

2.5 Scientific publications

EMERALD has published scientific publications in conferences and journals as part of the planned dissemination activities, as detailed in the Dissemination and Communication strategy (D6.2 [10]), and as documented in D6.4 [11] and D6.5 [12].

The Consortium adheres to the open science practices of the Horizon Europe program [13], by promoting and facilitating open cooperative work among project participants, with other related projects, and with the wider scientific community, as early and widely as possible, and following the principle of being "as open as possible, as closed as necessary". Research results published during the project are open access as long as the publication license authorises it and the commercial interests of the partners are aligned with it. Whenever possible, Consortium members favour the "green" open access route, whereby the published article or the final peer-reviewed manuscript are deposited in an online repository, before, at the same time as, or after publication, ensuring that the embargo period requested by certain publishers has elapsed..

The format used to ensure that data related to the scientific publications can be accessed is PDF format. The metadata used is adapted to that of the repository in which the paper is deposited and conforms to the format required by OpenAIRE [14], in order to facilitate indexing.

The consortium members may also use the Zenodo repository [15] for their joint publications. Some partners, such as TECNALIA, have developed their own OpenAIRE compatible repositories, to which researchers can upload their publications, which are then automatically indexed in OpenAIRE.

2.6 Project deliverables

EMERALD has generated a set of deliverables summarizing the main activities and outcomes of the project. Among these deliverables, only those of public dissemination level (marked as PU) are shared beyond the Consortium as open data (see Table 2). These deliverables are published in the **OpenAIRE** repository [14] (accessible via automatic publication once validated by the EC), on the project's **website**[4] and in the project's **Zenodo** repository[5]. Please note that the public deliverables of the project are only officially published on the website and Zenodo once they receive approval of the European Commission; in the meantime, they are labelled with a "draft" watermark.

All deliverables contain a set of keywords and a concise description intended to enhance their indexing and searching on search engines. These keyboards within each deliverable aim to

[4] https://www.emerald-he.eu
[5] https://zenodo.org/communities/emerald-he-project

CC BY-SA

emphasize the main topics covered in the document, whether it is a report or a software-related document.

Adhering to the principle of being "as open as possible as closed as necessary", we utilize Creative Commons licenses -specifically CC BY-SA 4.0 DEED[6]- for project outcomes that are not commercially IP-protected. These licenses ensure that the results are shared with minimal restrictions, while still requiring proper attribution to the authors or creators.

The audience for EMERALD's public deliverables spans a wide spectrum. It includes general audiences who are interested in the project's activities, as well as more specialized audiences such as developers and operators of multi-cloud applications. Additionally, those who want to understand the benefits of EMERALD through the experiences gained from the pilots are also part of the intended audience.

Table 2. Public deliverables in EMERALD

Number	Deliverable name	Short description	Leader	Type	Delivery date
D1.1 & D1.2	Data modelling and interaction mechanisms - v1/v2	Initial and final versions of the overview of data models and techniques used for creating and linking the data to evidence (annotation, etc)	**FABAR**	**Report**	**M9 & M18**
D1.3 & D1.4	EMERALD solution architecture - v1/v2	Initial and final versions of the description and design of the architecture of the EMERALD solution and underlying component integration.	**TECNALIA**	**Report**	**M12 & M24**
D1.5 & D1.6	DevOps methodology and CD/CI strategy for EMERALD - v1/v2	Initial and final versions of the description of the DevOps strategy and setting up of CD/CI environments	**TECNALIA**	**Report**	**M6 & M18**
D1.7 & D1.8 & D1.9	EMERALD Integrated solution – v1/v2/v3	Initial, intermediate and final integrated solution of the EMERALD audit suite	**TECNALIA**	**OTHER (SW)**	**M18 & M30 & M34**
D2.1	Graph Ontology for Evidence Storage	Report on the description of a uniform schema for storing and linking heterogenous data	**SCCH**	**Report**	**M9**
D2.2 & D2.3	Source Evidence Extractor – v1/v2	Interim and final evidence extraction from source code that can be integrated with the certification graph	**SCCH**	**OTHER (SW)**	**M12 & M24**
D2.4 & D2.5	AMOE – v1/v2	Interim and final evidence extraction from policy documents that can be integrated with the certification graph	**FABAR**	**OTHER (SW)**	**M12 & M24**

[6] https://creativecommons.org/licenses/by-sa/4.0/

(cc) BY-SA

Number	Deliverable name	Short description	Leader	Type	Delivery date
D2.6 & D2.7	ML model certification – v1/v2	Interim and final security and privacy preserving evidence that can be integrated with the certification graph	**FHG**	**OTHER (SW)**	**M12 & M24**
D2.8 & D2.9	Runtime evidence extractor – v1/v2	Interim and final evidence extraction from runtime data that can be integrated with the certification graph	**FHG**	**OTHER (SW)**	**M12 & M24**
D2.10 & D2.11	Certification Graph– v1/v2	Interim and final integration of the graph with semantically linked and combined evidence	**SCCH**	**OTHER (model)**	**M15 & M27**
D3.1 & D3.2	Evidence assessment and Certification – Concepts - v1/v2	Initial and final versions of the report on the requirements, design and integration of WP3 components	**FHG**	**Report**	**M9 & M18**
D3.3 & D3.4	Evidence assessment and Certification – Implementation - v1/v2	Interim and final versions of the implementation of the WP3 components.	**FHG**	**OTHER (SW)**	**M12 & M24**
D3.5 & D3.6	Evidence assessment and Certification – Integration - v1/v2	Interim and final integration of the WP3 components in the overall EMERALD system.	**FHG**	**OTHER (SW)**	**M15 & M27**
D4.1 & D4.2	Results of the UI-UX requirements analysis and the work processes – v1/v2	Initial and final versions of the report on the elicited UI-UX requirements from the target group. Work processes and workflows that should be covered with the user interface concept.	**KNOW**	**Report**	**M9 & M18**
D4.3 & D4.4	User interaction and user experience concept – v1/v2	Initial and final versions of the report on the developed mock-ups for the user interaction and user experience concept	**KNOW**	**Report**	**M12 & M24**
D5.1	Pilot definition, set-up & validation plan	Report on pilot set-up, validation plan of the user interaction concept, elicited requirements, and list of KPIs to measure the impact.	**FABAR**	**Report**	**M9**
D6.1	Project flyer and public website	First flier and project website, including project objectives and contact details.	**CNR**	**DEC**	**M4**
D6.2	Dissemination and Communication strategy	Report on how to target the different audience and which social media to use. Detailed plan for networking activities with external entities. Project	**CNR**	**Report**	**M6**

(cc) BY-SA

Number	Deliverable name	Short description	Leader	Type	Delivery date
		dissemination strategy to adopt throughout the project lifetime			
D6.3	Market analysis and applicability	Description of the preliminary analysis of the market, by inspecting solutions, trends, and initiatives in the fields relevant to the project. Description of the preliminary analysis of standardization activities.	**FABAR**	**Report**	**M6**
D6.4 & D6.5	Dissemination and Communication Report - v1/v2	Initial and final versions of the report on dissemination and communication activities followed during the reporting periods and the results obtained. Update of the dissemination and communication plan.	**CNR**	**Report**	**M18 & M36**
D7.2 & D7.3	Data Management Plan – v1/v2	Initial and final versions of the data management strategy of the project, and the established procedures and criteria to guarantee ethics and privacy in all project activities.	**TECNALIA**	**DMP**	**M6 & M36**

2.7 Other publications

EMERALD has also produced data for its dissemination, communication, networking, and exploitation activities under WP6, such as presentations, flyers, annual reports, etc. These data are freely available on the project website[7] and in the YouTube channel[8] in the form of accessible pdf or mp4 files. When IPR of foreground knowledge needs to be protected, the corresponding disclosures are published.

The list below describes the relevant data and the formats used for dissemination and communication materials:

- Press releases (pdf)
- Flyers (pdf)
- Annual summaries (pdf)
- Project presentations (pdf)
- EMERALD fragments (text, jpeg, pdf)
- EMERALD showcases (mp4)

In case videos and recordings containing personal data of participants are published, data subjects are asked to sign an informed consent before publication, as explained in section 6.1. In the event that any of the participants object to the public sharing of the data, the data may be processed to anonymise the information if possible or kept private if not.

[7] http://www.emerald-he.eu
[8] https://www.youtube.com/@emerald-he-project

CC BY-SA

2.8 Data used during the CI/CD process

During continuous integration and continuous delivery (CI/CD) processes there are some stages in which data may be required:

- During the building of a component, data may be needed to setup the component.
- During the verification processes, data may also be required to perform some tests.
- Finally, during the deployment processes, data may be needed to configure the component for the specific environment.

The data to setup a component during the build phase should not be related to any of the pilots, and in the odd case that it is related to any of them it must not contain sensitive information. For testing purposes, partners may use synthetic, fictitious or 'persona' data, but under no circumstances may they use real data, anonymized or not, from the pilots.

During the automatic verification processes of components prior to their deployment in the production and pilot environments, it may be necessary to use verification data. This data will consist of several data sets that shall be similar to the data that will be used by the components in the real environment. These data sets must not be real or include sensitive data in order to be used in the process. Therefore, they are no relevant from the perspective of the data management plan.

Finally, data will also be required during the deployment process. Unlike the other stages (build and verification), this information will most likely include sensitive information, e.g., passwords needed for deployment in the pilot. Credentials or secrets, owned by the pilot concerned, must never be committed to the GitLab, instead, will be managed using environment-variable injection practices.

2.9 Data stored in the Repository of metrics

As part of the project deliverables, a structured collection of security metrics for continuous certification of cloud services has been developed and is publicly available[9]. This Collaborative Metric Repository contains:

- Metric definitions (YAML files), describing measurable security properties for cloud resources, including target values and comparison operators.
- Automated evaluation policies (Rego/OPA files), enabling programmatic assessment of compliance.
- An ontology, harmonising resource types and security properties across cloud providers and certification schemes.
- Sample resource files (JSON) for testing and demonstrating metric evaluation.

The data is published under the Apache 2.0 license, which permits unrestricted use, redistribution, and adaptation with attribution. No access restrictions apply. The repository is maintained as a living resource and is updated as new metrics are developed within the project.

[9] https://github.com/Cybersecurity-Certification-Hub/security-metrics

CC BY-SA

3 FAIR Data

This section presents the measures to ensure the compliance of the project data with the FAIR data principles as defined by the EC guidelines [16], which refer to the feasibility and appropriateness of making data findable, openly accessible, interoperable and reusable in the context of the project.

- **Findability**: Open (meta)data should be easily searchable and locatable. This involves including identifiers, keywords, version numbers, and metadata to maximize the possibilities of finding and reusing the data.
- **Accessibility**: The openness of (meta)data must be guaranteed, and the means of access clarified and enabled. In addition to details about the repository where the data will be deposited, the methods and protocols for access need to be specified, along with any potential restrictions.
- **Interoperability**: Refers to the ability of (meta)data to integrate or work seamlessly with other (meta)data and tools. Therefore, it is essential to clarify the vocabularies, standards, formats, and methodologies that will be used to enable (meta)data exchange, reuse, and interoperation.
- **Reusability**: (Meta)data should be usable in other contexts, including accurate and relevant attributes, clear and accessible (meta)data usage licenses, compliance with domain-relevant standards, and identification of provenance.

3.1 Making data findable, including provisions for metadata

EMERALD adheres to standard naming conventions, incorporates relevant search keywords, implements version numbering, and utilizes metadata to ensure easy discovery and identification of data.

The Digital Object Identifier (DOI) serves as the standard identification mechanism, providing persistent and unique identifiers for the data. Rich metadata plays a crucial role in enabling other researchers to locate and reuse the data effectively.

To characterize the data, metadata can encompass various types of information, including DOI, title, creation date, publication date, version number, author, publisher, copyright, license, keywords, grant agreement number, project acronym, and data format. Keywords are provided to optimize search and re-use, and they can also be part of the metadata.

For open scientific publications, the following bibliographic metadata is employed to identify and facilitate access to them:

- Funding grant: EC Horizon Europe
- EMERALD, Grant Agreement No. 101120688
- Publication date, length of any applicable embargo period, and a persistent identifier

For public deliverables, the naming convention to be used is "*EMERALD_Dx.y Deliverable name _version.pdf*". All deliverables include a set of keywords and a brief description intended to facilitate the indexing, searching of the deliverables in search engines and re-using. The keyboards in each deliverable aim to stress the main topics addressed in the document. The public deliverables will be stored at the hosting provider of the partner CNR (responsible for WP6 "Dissemination, exploitation and communication") for at least three years beyond the end of the project and, for long-term storage, in the Zenodo repository (see Section 3.2.1).

(cc) BY-SA

3.2 Making data openly accessible

The EMERALD project is committed to follow the European Commission's guidelines for open access to research data in all funded Horizon Europe projects [17]. Therefore, the project relies on widely used open repositories to deposit the project generated datasets.

3.2.1 Repositories

The project has set up a **private document repository** where all the project-related documents are being collected and shared (see Figure 3). is provided by the Fabasoft partner and hosted in the Fabasoft Cloud. The repository provides a role-based access management of the contents for the safe storage of the project outcomes. It will be available at least 1 year after the project has finished. FABAR and TECNALIA can then jointly agree on "keeping it alive" for a longer period and/ or moving data as backups and documentation to other systems.

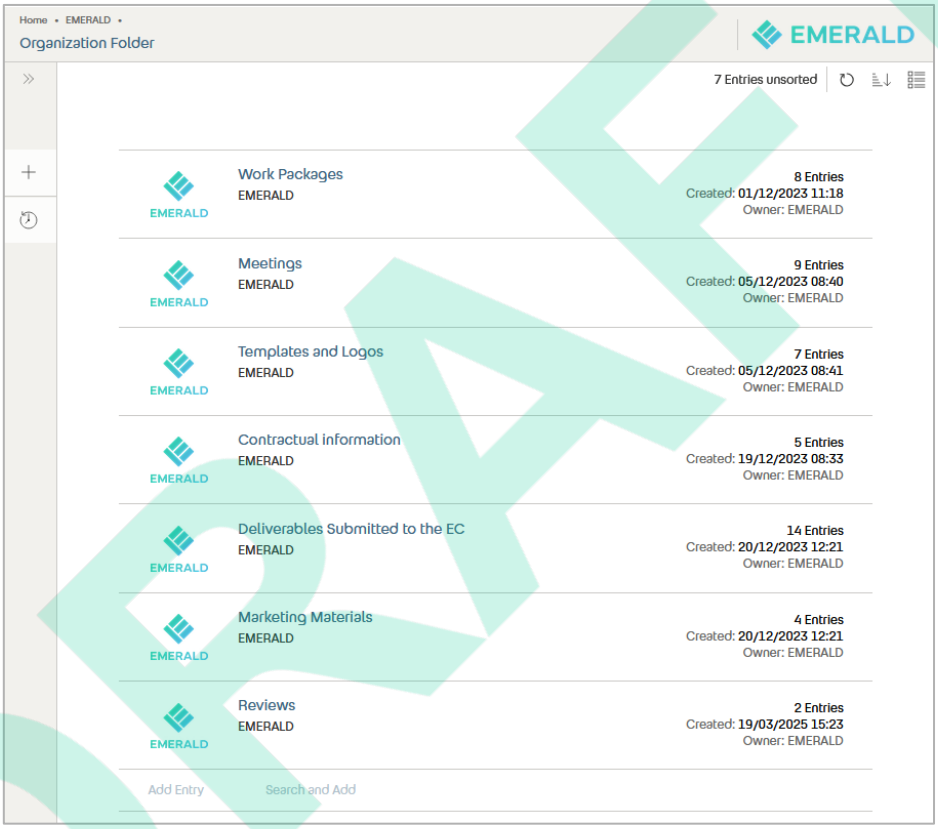

Figure 3. EMERALD private document repository in Fabasoft Cloud

A **private repository on Gitlab** has been set up by TECNALIA to facilitate the development of EMERALD's software components. This repository is used to manage the source code of the EMERALD components, so that several people can work on the implementation, whilst maintaining a track of the changes; to set up the DevOps pipelines and instruments to implement the Continuous Integration and Continuous Deployment (CI/CD) strategy; to manage EMERALD requirements, which are tracked in a specific repository via issues; and to support architecture modelling, maintaining the various diagrams as code and enabling their visualization via web. Access to this repository will be available for the duration of the project and will remain available for at least 5 years after the end of the project.

DRAFT

CC BY-SA

In addition, the open-source software developed in the project is available in a **public GitLab repository**[10] provided by TECNALIA (see Figure 4). Access to this repository will be available for the duration of the project and will remain available for at least 5 years after the end of the project. The repository's web address is published in the "Open Source" section of the project's official website.

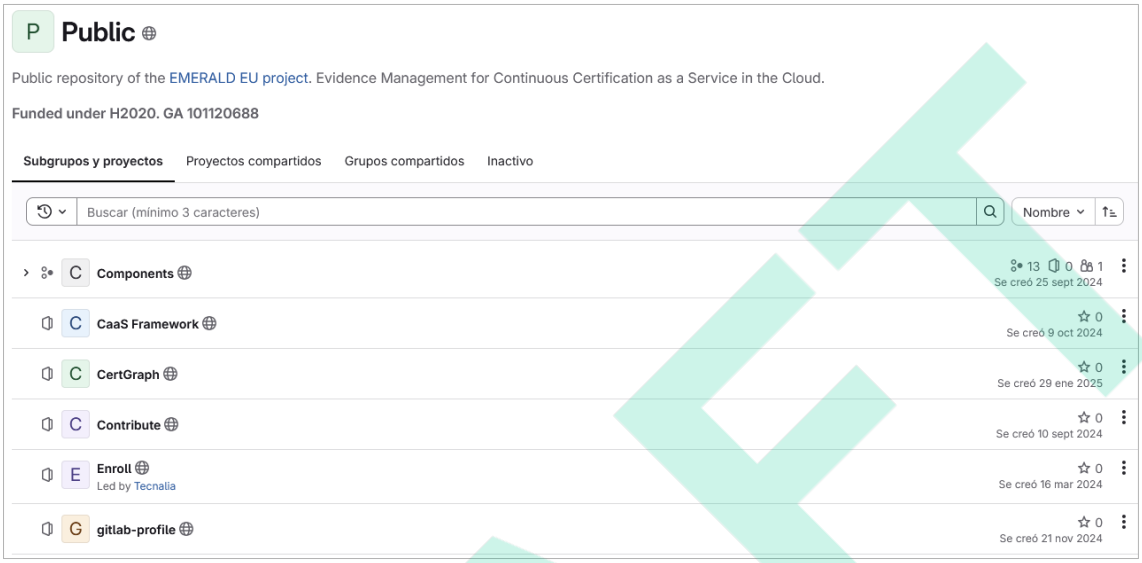

Figure 4. Public Gitlab repository for EMERALD

The EMERALD project is committed to adhering to the European Commission's guidelines for open access in all funded Horizon Europe projects [17], so our plan involves relying on a widely used open repository to store the data generated by the project. In particular, the **open repository Zenodo** [15] is the preferred choice within the Consortium. Zenodo is securely hosted and operated by CERN and integrates seamlessly with the OpenAIRE platform where public data from the EMERALD project will be stored. Data are licensed under CC (Creative Commons "No Rights Reserved"). It is important to note that uploading results to Zenodo does not alter the ownership rights of the content. Zenodo also provides a connector to GitHub that supports open collaboration for source code and version control for all types of data.

EMERALD has created a dedicated **Zenodo community**[11] that will serve as the main repository for the open data resulting from the project (see Figure 5). All public deliverables produced during the project are uploaded to Zenodo for long term storage and open access. Furthermore, Zenodo also provides access to the project's outreach materials, such as leaflets and annual summaries.

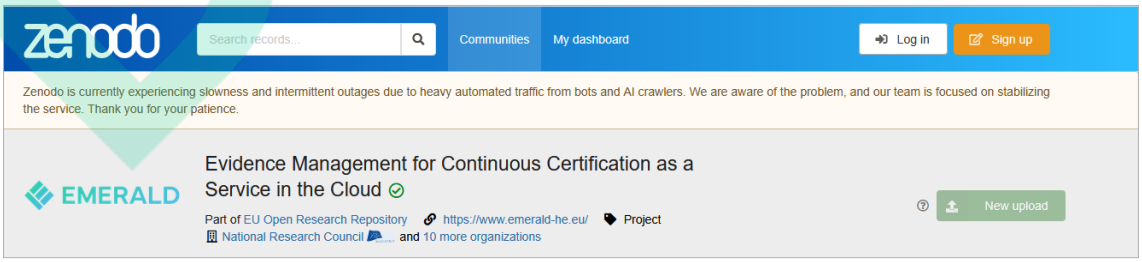

Figure 5. Zenodo community for EMERALD

[10] https://git.code.tecnalia.dev/emerald/public
[11] https://zenodo.org/communities/emerald-he-project

CC BY-SA

For scientific publications, the project will give priority to "green" open access, where the published article or final peer-reviewed manuscript is deposited in an **online repository** (either before, at the same time or after publication), to ensure that any embargo period requested by specific publishers has lapsed. Each partner will assume responsibility for providing an **open access option** for every scientific article and paper it publishes and will ensure that the publication is made openly accessible within a maximum of six months. In addition, all open publications and public project deliverables will be made accessible via the project website[12].

Finally, the project has been involved in the implementation of the **Collaborative Metric Repository**[13], which collects security metrics that can be used for continuous certification of metrics, and is a collaborative outcome of EMERALD and other projects participating in the **EC3** (European Cluster for Cybersecurity Certification)[14]. The repository is structured as follows.

- *metrics*: This folder contains the metrics. It is organised first by domain, and each domain contains several folders of metrics (see Figure 6).
- *ontology*: This folder holds the ontology on which the metric descriptions are based. The ontology describes which resources a system may consist of, which security features they may offer, etc. The metric descriptions are based on the terms in the ontology.
- *resources*: This folder contains sample files with structured information about resources, based on the ontology, in JSON format.
- *templates*: This folder contains a metric template that can be used to create new metrics.

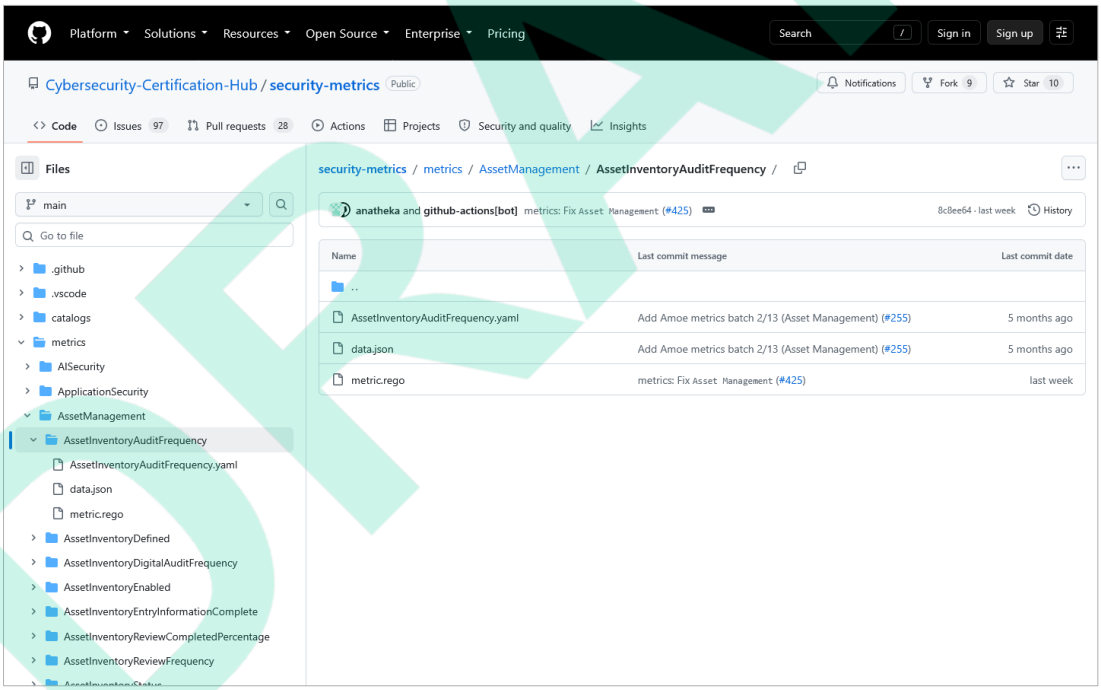

Figure 6. EMERALD metrics in the Collaborative Metric Repository

3.2.2 Metadata

As explained in section 3.2.1 **Zenodo** is the preferred long term storage option for EMERALD. All uploaded results in Zenodo are structured using metadata such as names, keywords, date,

[12] https://www.emerald-he.eu
[13] https://github.com/Cybersecurity-Certification-Hub/security-metrics
[14] EC3 web site: https://cybersecuritycertcluster.eu/

(cc) BY-SA

location, document type and licence information. These metadata provides comprehensive information, is stored internally in JSON-format according to a defined JSON schema and provides exporting mechanisms to several standard formats such as MARCXML, Dublin Core, and DataCite Metadata Schema. Metadata is openly available and contain information that enables the user to access the data.

In addition, EMERALD partners use a **Gitlab public repository** provided by TECNALIA to gather all the software declared as Open Source (see section 3.2.1). The metadata related with each software file is mainly the author's username, the production date and commit messages made by the author. More general metadata, such as module description, functionality, usage and installation instructions, is also available in the readme files of each component repository. Metadata will be available for the project duration and will remain available for at least 5 years after the end of the project.

The metrics of the **Collaborative Metric Repository** are also described by metadata such as id, name, description, version, comments, interval, operator and targetValue. The repository also includes an ontology[15], which has been created using *Protegé* and is published in the online Protegé platform *webprotege*. The ontology was developed to harmonize evidence gathering and assessment across certifications, cloud vendors, and resource types. It includes several taxonomies, including a taxonomy of cloud resources and a taxonomy of security properties.

3.2.3 Data related to the pilots

The data managed by the pilots will not be openly accessible. All data collected, stored and processed by the pilots will be treated as strictly confidential and will be kept for a specific period of time no longer than necessary to achieve the objectives of the pilot and to validate the project objectives. The results of the validation tasks to be carried out by the pilots will be anonymized and aggregated in the final WP5 deliverable (D5.6 [18]).

For each EMERALD pilot, we have analysed the location of the back-end data, who is responsible for managing it and who has access to it.

3.2.3.1 Pilot 1: Certification of public Cloud Services – IONOS

Where will the back-end data be stored for the pilot?

The backend data for the pilot implementation will be stored within the IONOS public cloud service infrastructure.

Who will be in charge of managing data?

Data management responsibilities for the pilot will be handled by the project team within IONOS.

Who will have access to the data?

Access to the data will be restricted to authorized project team members involved in the pilot implementation, this can include other EMERALD partners if required.

3.2.3.2 Pilot 2: Certification of public Cloud Services – CloudFerro

Where will the back-end data be stored for the pilot?

Test environments for IaaS and PaaS will be deployed on one of CloudFerro's public cloud.

[15] https://github.com/Cybersecurity-Certification-Hub/security-metrics/blob/main/README.md

CC BY-SA

Who will be in charge of managing the data?

Pilot 2 data will be managed by CloudFerro.

Who will have access to the data?

CloudFerro will have access to Pilot 2 data, other EMERALD partners may have access if necessary.

3.2.3.3 Pilot 3: Certification of public Cloud Services – Fabasoft

Where will the backend data be stored for the pilot?

- app.telemetry logs: The log data is stored on Fabasoft's infrastructure and provided to EMERALD through an API plug-in that makes use of industry standards to communicate the result.
- Objects in Fabasoft Cloud/Fabasoft Folio: The object data are stored on Fabasoft's infrastructure and can be collected through an API plug-in that makes use of industry standards to communicate the result.
- Submitted evidence as well as documents uploaded to *AMOE* will be stored in the respective EMERALD framework components and database deployments.

Who will be in charge of managing the data?

For all data connected to the Pilot 3 and hosted on Fabasoft's infrastructure (log data, object data, manual controls), Fabasoft will be managing it.

Data that is provided in direct connection to Pilot 3 and is no longer stored on Fabasoft's systems (i.e., data that is accessed through the API and policy documents) and therefore leave Fabasoft's premises, is under the responsibility of the respective partner in charge of the component that processes the data.

Who will have access to the data?

Fabasoft's data is accessible through a common interface after successful authentication. Those components that need to collect information to fulfil their purpose and are dependent on Fabasoft's data will get access only and after approval of the Security Advisory Board (SAB) for the purpose of Pilot 3 and Pilot 4 activities to fulfil the project goals of EMERALD. Partners receiving this data are responsible for proper data handling.

3.2.3.4 Pilot 4: Certification of hybrid cloud-edge environments for the financial sector

Where will the backend data be stored for the pilot?

For Pilot 4, the back-end data will be under the security perimeter of the CSPs that compose the use case architecture. Pilot owners will ensure that sensitive data will not be stored during pilot activities execution or used outside of the pilot testing.

Who will be in charge of managing the data?

CaixaBank, which manages the audit scope, the policy documents and the mapping of metrics to controls, and OpenNebula, which manages the test environment and its evidence collector.

(cc) BY-SA

Who will have access to the data?

Technical staff of CaixaBank and OpenNebula, and the auditors of DNV Cyber during the audit.

3.3 Making data interoperable

To enhance the re-use of open data from external sources and to enable wider utilization of EMERALD's open data beyond the project, EMERALD actively promotes and adopts generic format standards that are compatible with widely available software programs.

The consistent use of common terminology across all data types plays a crucial role in facilitating data interoperability. As part of this effort, the EMERALD project has developed a common taxonomy and ontology for resource types, giving rise to the EMERALD ***CertGraph*** [19]. *CertGraph* is a graph-based knowledge structure that provides a common representation of certification information and enables evidence to be linked, combined and traced back to its original source, supporting a more consistent approach to automated assessment.

EMERALD also has developed ***AIOX***, an AI-assisted tool that helps experts extend the certification ontology when new security metrics are introduced [19]. The tool identifies concepts that are not yet represented and suggests their relationships and placement, while keeping human experts responsible for validating the proposed changes.

3.4 Increase data re-use

The European Commission characterizes Open Access as the availability of scientific information online, accessible to users at no cost and suitable for reuse. The EMERALD project actively contributes to this approach, widely recognized for its positive impact on science, innovation, and efficiency, through the act of depositing data into an open access repository such as Zenodo (see section 3.2.1), to make it more reusable and accessible.

Wherever possible, data shall be licensed for open access. Where open access is not possible, alternative licensing options are explored. In any case, EMERALD adheres to the principle of being "as open as possible, as closed as necessary".

Data owners use licences to specify access permissions to their data. It is important to note that in most cases, as with Creative Commons licences, these licences relate primarily to copyright and related matters, excluding the protection of personal information.

DRAFT

CC BY-SA

4 Other Research Outputs

As shown in Table 1, some of the project's public deliverables (classified as "OTHER") will not consist solely of a report but will include other types of outputs. These additional outputs may be software, models or algorithms.

As indicated in the EMERALD DoA [5], the software developed in the project will be published in a public open-source repository under OSS license, in accordance with the guiding principles of the European Commission's Open-Source Software Strategy 2020-2023. Some components coming from previous projects such as MEDINA are provided under existing open-source licenses, licences that will be maintained. Specific details concerning the OSS use are being thoroughly addressed in the EMERALD Consortium Agreement [20].

The repository used to publish the EMERALD open-source code is a GitLab instance[16] provided by TECNALIA. Open-source versions of the different release of EMERALD components will be accessible, findable and reusable by any developer interested in the EMERALD results. For each software component, a readme file and a technical specifications document are available. In addition, a docker script is provided that will allow users to deploy the container in any desired infrastructure.

Finally, the EMERALD consortium will seek to update, maintain, and extend the open-source code released in EMERALD also after the end of the project, and will seek to engage the community of developers who will contribute to the update, maintenance and release of new functionalities. The Project Coordinator, TECNALIA, will ensure that the public Gitlab repository is available for at least 5 years after the project termination, and will transfer it to existing open-source projects and communities such as Zenodo to assure long-term preservation.

DRAFT

[16] https://git.code.tecnalia.dev/emerald/public

CC BY-SA

5 Allocation of Resources

The Project Coordinator, TECNALIA, will be responsible for appointing the EMERALD project DPO (Data Protection Officer), who is the primary contact for the personal data protection issues in the project and will oversee the managing of the data protection issues together with the partners' appointed DPOs.

The partner FABAR assumes responsibility for the maintenance of the project's document repository (Fabasoft Cloud) and the partner TECNALIA assumes responsibility for the maintenance of the project's open-source software repository and the private development repository (Tecnalia GitLab). These repositories are free of charge for the project. In addition, each partner assumes responsibility for ensuring the accessibility of the data it produces.

The Project Coordinator, TECNALIA, and the Technical Coordinator, Fraunhofer, will monitor the quality of all scientific data results. Some EMERALD partners (FABAR, CXBTECH, IONOS, CF, ONS and NIXU/DNV) have appointed a member for the SAB (Security Advisory Board) that will support the Project Coordinator in the management of security aspects, as explained in section 6.2.

As expressed above, open access repositories will be favoured, so EMERALD does not foresee any additional needs beyond the duration of the project to manage data or make it fair.

DRAFT

CC BY-SA

6 Data Security

As stated in the DoA [5], the EMERALD project does not use classified background information, nor does it produce classified foreground information.

The project has set up two private repositories, as explained in section 3.2.1:

- A private document repository where all the project-related documents are collected and shared. It is provided by the partner FABAR and hosted in the Fabasoft Cloud. The repository provides a role-based access management of the contents for the safe storage of the project outcomes.
- A private Gitlab repository to support the development of EMERALD software components, provided by TECNALIA. The access to GitLab is account-based and includes two-factor authentication (2FA) which provides an additional level of security (for others to access your account, they would need your username and password and access to your second factor of authentication). The repository access is only allowed to project members, after individual request. Also, the accounts provide a role-based access, and the role with the less privileges needed is assigned by default.

The development of the EMERALD components adheres to the principles of security-by-design and privacy-by-design that allow security and auditing procedures to remain consistent across multiple environments. In addition, as part of the EMERALD Framework, the following functionalities have been included in the architecture:

- Access Management, which grants authorized users the right to use a service, while preventing non-authorized users from accessing it.
- Credentials Management, to manage credential information such as usernames and passwords.

Keycloak[17] has been chosen as open-source Identity and Access Management (IAM) solution for the project to manage users and roles, and to secure the communication among components through role-based access control (RBAC).

The two relevant procedures established by EMERALD in relation to the protection of personal data processed in the project are described below.

6.1 Personal data protection management

As required in the Grant Agreement and the Consortium Agreement, the EMERALD partners are committed to process personal data in compliance with applicable data protection laws (in particular, the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data [21]).

Confidentiality clauses in the Grant Agreement and the Consortium Agreement ensure that all beneficiaries are required to limit the access of their staff to personal data, unless it is strictly necessary for the implementation, management or monitoring of the project activities. In particular, beneficiaries shall, where necessary, conclude a separate data processing, data sharing and/or joint controller agreement before any data processing or data sharing takes place.

The technical research approach of the EMERALD project does not require the processing of personal data. However, during the implementation of project activities some personal data

[17] https://www.keycloak.org/

CC BY-SA

may be required. Whenever personal data is about to be collected the data subjects involved shall be duly informed and requested to sign a consent form, and the processing shall be limited to the strictly needed amount and types.

During the organization of project meetings some personal data of the attendants may be required, as well as group pictures or recordings for the preparation of the minutes, which may contain images and voiced of the presenters, subject to the signature of a general informed consent, cf. *APPENDIX A: General Informed Consent Form*.

In the context of WP4 activities, some interviews will be performed with people involved in the pilots to elicit the requirements for the development of the integrated EMERALD UI. These interviews, which may contain images and voices of the interviewers, will be recorded subject to informed consent, cf. *APPENDIX B: Information for Conducting Interviews in WP4*.

In addition, as part of the WP5 validation activities, some videos showing the validation of the EMERALD solution in the pilots could be recorded in mp4 format, which may contain images and voices of the presenters, subject to informed consent.

Finally, in the context of WP6 dissemination activities, all stakeholders will be asked to provide their consent prior to their participation in EMERALD events, webinars, and meetings. The Consent form will ensure that participants are informed and agree to the processing of specific personal data including: first and last names, email addresses, country, and type of organization that will be stored in an xlsx file, and images captured in jpg format. Recordings of the webinars (in mp4 format) may contain images and voices of both presenters and attendees. Participants' consent will be obtained for the use of these recordings.

6.2 Security management

The security management in EMERALD refers to the implementation of security measures to prevent the disclosure of security sensitive information in the project deliverables and data.

As stated in the DoA [5], the EMERALD project does not use classified background information or produce classified foreground information, and no security issues have been identified in the Security Self-assessment conducted in Part A, under section "4. Ethics and Security".

However, specific measures have been outlined in the Consortium Agreement and will be continuously monitored and updated throughout the project to ensure secure handling of project information and results. All actions will be subject to the supervision of the Project Coordinator and the Technical Manager, who will monitor all project deliverables to prevent any security issue arising from them. The Consortium has also set up a private document repository supported by the Fabasoft Cloud infrastructure for the safe storage of project outcomes and role-based access management of the contents has been implemented.

A Security Advisory Board (SAB) has also been set up in EMERALD for the protection of security sensitive information in certain SEN-type deliverables, as detailed in D7.1 [22]. The mission of the SAB is to assist the Project Coordinator in the design and implementation of relevant measures to effectively manage, mitigate and respond to security concerns as they arise. Every partner involved in the EMERALD pilots has appointed a person with an excellent background in cybersecurity to be part of the SAB.

The SAB members will review deliverables that may contain security sensitive information, with the objective of identifying any aspects that need to be protected. In the event that the review uncovers some security-sensitive information, the SAB will promptly collaborate with the editors to establish measures and guidelines to prevent inadvertent disclosure of information.

CC BY-SA

Finally, to ensure that the entire Consortium is kept abreast of the developments in security management, the Project Coordinator will present a summary of security oversight activities, including decisions made and results achieved, at the General Assembly meetings. In addition, guidance will be provided at these meeting to members on how to safeguard security-sensitive information from inadvertent disclosure.

DRAFT

CC BY-SA

7 Ethics

No harmful ethics impact is expected from the project activities on the environment and society, as stated in the DoA [5]. Some EMERALD components use AI technologies for their development, but these technologies do not involve the processing of personal data:

- ***Mapping Assistant for Regulations with Intelligence* (*MARI*)**: This component uses state-of-the-art NLP tools to check the similarity of natural language texts. These texts express cloud security requirements and security metrics, and their similarity will be exploited to associate metrics with requirements, and to establish which requirements of different certification schemes are similar to each other. Since the analysed texts do not concern personal or sensitive data, no ethical issues emerge in the use of Artificial Intelligence to implement the *MARI* component.
- ***Assessment and Management of Organizational Evidence* (*AMOE*)**: This component extracts information on prepared, non-confidential policy data and texts. The only nominative and personally identifiable data could be considered in the application of the *AMOE* software within the pilots, but this will be substituted by creating fake data for the roles involved in the certification process. This is needed because the name of the person verifying and approving a policy document has to be verifiable for the auditor to examine the application and update process of such documents. To summarize: training and test data for *AMOE* does not work with personal data.
- ***AI Security Evidence Collector* (*AI-SEC*)**: This component is used to extract evidence on the security and privacy evaluation of Machine Learning models used in Cloud applications. It is connected to the application of ML methods for source code analysis and for referencing evidence to the *CerthGraph*. Additionally, its main focus is to elaborate on the PoC for scaling the EMERALD approach to AI system certification – i.e., it deals with the question whether or not other AI systems violate privacy requirements. *AI-SEC* does not deal with personal data and uses AI models and public datasets.

In conclusion, no ethical issues related to human rights and values have been identified in the Ethics Self-assessment that was performed in Part A, under section “4. Ethics and Security”.

In any case, the Project Coordinator will monitor the ethical aspects of all project activities and deliverables, ensuring compliance with relevant ethical standards, in particular with the General Data Protection Regulation 2016/679 and the EU policy actions and guidelines relevant to AI (e.g., the Ethics Guidelines for Trustworthy AI, the AI Whitepaper, the EU Security Strategy and the Data Strategy). In addition, all project participants will work closely with the Project Coordinator, promptly reporting any instances of misconduct, misuse, or other issues that may arise during the project research activities.

DRAFT

CC BY-SA

8 Conclusions

This document presents the final version of the Data Management Plan (DMP) for EMERALD, which serves as a comprehensive guide for project partners to meet data management requirements.

The DMP describes all types of data handled by the project, including interviews and focus groups data, pilot data, scientific publications, public deliverables, other publications, data used during the CI/CD process and data stored in the repository of metrics. It outlines procedures and criteria to ensure that data generated by the project adhere to FAIR (Findable, Accessible, Interoperable and Reusable) principles and sets out the procedures for ensuring compliance with data protection legislation (i.e. the General Data Protection Regulation) in data management.

To enhance impact and promote open science, EMERALD adheres to the principle of being "as open as possible, as closed as necessary". Specifically, certified open data repositories such as Zenodo, OpenAIRE and GitLab are used to deposit the project's open data; whenever feasible, the green open access model for scientific publications is favoured, and all open data generated within the project is published or referenced on the project website.

The DMP has been an evolving plan throughout the course of the project and has been reviewed periodically. The final version of the plan, delivered at the end of the project (M36), is based on and refines the previous initial version from M6, which is documented in deliverable D7.2 [1].

DRAFT

CC BY-SA

9 Referencias

[1] EMERALD Consortium, “D7.2 Data Management Plan–v2,” 2024.

[2] European Commission, “Data Management Plan (HE),” April 2022. [Online]. Available: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx. [Accessed October 2026].

[3] MEDINA Consortium, “H2020 MEDINA project (GA 952633),” [Online]. Available: www.medina-project.eu. [Accessed October 2026].

[4] MEDINA Consortium, “D5.5 MEDINA integrated solution-v3,” 2023. [Online]. Available: https://medina-project.eu/wp-content/uploads/2023/09/MEDINA_D5.5_MEDINA-integrated-solution-v3_v1.0.pdf. [Accessed October 2026].

[5] EMERALD Consortium, “EMERALD - Annex 1 - Description of Action - GA 101120688,” 2022.

[6] EMERALD Consortium, “D1.9 EMERALD solution architecture-v3,” 2026.

[7] EMERALD Consortium, “D4.1 Results of the UI-UX requirements analysis and the work processes–v1,” 2024.

[8] EMERALD Consortium, “D4.2 Results of the UI-UX requirements analysis and the work processes – v2,” 2025.

[9] EMERALD Consortium, “D5.6 Evaluation report and impact analysis”.

[10] EMERALD Consortium, “D6.2 Dissemination and Communication Strategy,” 2024.

[11] EMERALD Consortium, “D6.4 Dissemination and Communication Report - v1,” 2025.

[12] EMERALD Consortium, “D6.5 Dissemination and Communication Report-v2,” 2026.

[13] European Commission, “Open access in Horizon Europe,” [Online]. Available: https://rea.ec.europa.eu/open-science_en. [Accessed October 2026].

[14] “OpenAIRE,” [Online]. Available: https://www.openaire.eu/. [Accessed October 2026].

[15] “Zenodo,” [Online]. Available: http://www.zenodo.org. [Accessed October 2026].

[16] M. D. M. A. I. e. a. Wilkinson, “FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018,” 2016. [Online]. Available: https://doi.org/10.1038/sdata.2016.18. [Accessed October 2026].

[17] European Commission, “HE Programme Guide,” [Online]. Available: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf. [Accessed October 2026].

DRAFT

(cc) BY-SA

[18] EMERALD Consortium, «D5.6 Evaluation report and impact analysis,» 2026.

[19] EMERALD Consortium, “D2.11 Certification Graph -v2,” 2026.

[20] EMERALD Consortium, “EMERALD Consortium Agreement,” 2024.

[21] European Commission, “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (‘GDPR’),” 2016.

[22] EMERALD Consortium, “D7.1 Project Manual and Quality Plan,” 2024.

DRAFT

(cc) BY-SA

APPENDIX A: General Informed Consent Form

EMERALD

AUTHORISATION OF THE USE OF IMAGES AND/OR VOICE

The undersigned, Mr./Mrs. _____, of legal age and with ID number _____ (hereinafter, the "**INTERESTED PARTY**") will voluntarily participate in the project entitled "Evidence Management for Continuous Certification as a Service in the Cloud" in short "EMERALD" (hereinafter, the "**PROJECT**"), in which the following entities participate:

FUNDACIÓN TECNALIA RESEARCH & INNOVATION (Project Coordinator), organised and existing under the laws of Spain, with its registered office at Parque Científico y Tecnológico de Gipuzkoa, Mikeletegi Pasealekua 2, E-20009 Donostia-San Sebastián, Spain, and VAT number G-48975767,

FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV (Participant), organised and existing under the laws of Germany with its registered office at Hansastraße 27c, 80686 München, and VAT number DE129515865;

FABASOFT R&D GMBH (Participant), organised and existing under the laws of Austria, with its registered office at Honauerstraße 4, 4020 Linz, Austria, and VAT number ATU63399566;

CONSIGLIO NAZIONALE DELLE RICERCHE (Participant), organised and existing under the laws of Italy, with its registered office at Piazzale Aldo Moro 7 00185, Roma, Italy, and VAT number IT02118311006;

SOFTWARE COMPETENCE CENTER HAGENBERG GMBH (Participant), organised and existing under the laws of Austria, with its registered office at SOFTWAREPARK 32a, HAGENBERG, 4232, Austria, and VAT number ATU 48056909;

KNOW-CENTER GMBH RESEARCH CENTER FOR DATA-DRIVEN BUSINESS & BIG DATA ANALYTICS (Participant), organised and existing under the laws of Austria, with its registered office at Sandgasse 36/4, 8010 Graz, Austria, and VAT number ATU50367703;

CAIXABANK SA (Participant), organised and existing under the laws of Spain, with its registered office at Calle Pintor Sorolla, número 2-4 (46002), Valencia, Spain, and VAT number ESA08663619;

IONOS SE (Participant), organised and existing under the laws of Germany, with its registered office at Elgendorfer Straße 57 56410 Montabaur, Germany, and VAT number;

CLOUDFERRO SP ZOO (Participant), organised and existing under the laws of Poland, with its registered office at Nowogrodzka 31, 00-511 Warsaw, and VAT number 7010468205;

OPENNEBULA SYSTEMS SL (Participant), organised and existing under the laws of Spain, with its registered office at PASEO DEL CLUB DEPORTIVO 1 - EDIFICIO 13 PARQUE EMPRESARIAL LA FINCA, POZUELO DE ALARCÓN 28223, Spain, and VAT number ESB85929941;

© EMERALD Consortium Contract No. GA 101120688 Page 1 of 2
www.emerald-he.eu
(cc) BY-SA

DRAFT

NIXU OYJ (Participant), organised and existing under the laws of Finland, with its registered office at Keilaranta 15 B, 02150 Espoo, Finland, and VAT number FI07218117;

hereinafter, jointly referred to as the "**EMERALD Consortium**".

In this context, the INTERESTED PARTY hereby makes the following **STATEMENTS**:

I. That he/she is aware of and expressly consents to the recording of his/her image and/or voice, through the relevant technical means, by the EMERALD Consortium during his/her participation in the PROJECT, all for the purpose of promoting and disseminating the PROJECT or other similar events organised by the aforementioned EMERALD Consortium.

II. That for the aforementioned purposes, he/she expressly authorises the EMERALD Consortium to (i) use his/her image and/or voice; and (ii) reproduce them through all technical means currently known, in any medium, tangible or intangible, and in particular by reproducing them on the web pages and corporate profiles on social networks that the EMERALD Consortium maintains accessible through the Internet, and/or through other media that, in different media, tangible or intangible, said entity uses to publicise its activities and professional services, all for the purpose of promoting and disseminating the development of the PROJECT and other similar events. Said authorisation is granted free of charge and shall, in any case, be revocable under the terms established by law.

III. That, by signing this document, he/she consents to the processing of his/her personal data in the following terms:

Responsible	EMERALD Consortium
Purpose	Promotion and dissemination of the PROJECT and/or other similar events.
Rights	You have the right to withdraw your consent at any time, oppose the processing, access, rectify and delete the data, as well as to exercise other rights, by sending an e-mail to dpo@tecnalia.com.

Name and surname:

Place and date:

Signature:

DRAFT

APPENDIX B: Information for Conducting Interviews in WP4

Please find below the contents of the Participant Information Sheet, the Informed Consent, and the Data Protection Information used in WP4 for conducting the interviews.

Participant Information Sheet

Participant Information Sheet

Introduction

You are invited to participate in an interview study related to the EMERALD Project. Before deciding on whether you want to participate or not, please read the information below. Please ask the researcher all the questions you may have so you are completely sure that you understand all the proceedings of the study. The contact details are provided at the end of this information sheet.

Purpose of the study

EMERALD is an HEU Project (GA no.: 101120688) with the objective to pave the road towards Certification-as-a-Service (CaaS) for continuous certification of harmonized cybersecurity schemas like the EUCS. This interview is conducted within WP4 – User Interaction and User Experience Development of the EMERALD Project. The goal of this interview is to elicit requirements of [auditors/chief information security managers/compliance managers] necessary for developing the integrated EMERALD UI.

In more detail, our goal is to elicit in-depth insights about your work as [auditors/chief information security managers/compliance managers] in relation to continuous cloud auditing processes. Therefore, we are conducting a series of interviews aiming at getting …

- … a good understanding of your work in general,
- … your activities and tasks in the cloud computing systems certification process,
- … insights on how EMERALD could support your working activities,
- … insights about the expectations towards the EMERALD UI,
- … insights about existing pain points,
- (… and if you have been in the MEDINA project, what went good or not so good in MEDINA, and what could be done better or different in EMERALD)

Your participation in the study

You are invited to participate in this study on a voluntary basis and you are free to withdraw from the study at any time without providing any reason for doing so. If you agree to participate in this interview, you give us permission to:

- Collect information from you;
- Share information (only answers you provide without any personal information) with the people of the project;
- Conduct the study;
- Use this information in the analysis and for publication.

Benefits of the participation

It is likely that you might not receive any direct personal benefit for your participation in this interview besides possibly learning more about the EMERALD project in general. However, by participating you will make a substantial contribution to the success of the EMERALD project, as we need your expertise for developing a good and easy-to-use EMERALD UI/UX that supports you during your work.

CC BY-SA

Disadvantages and/or risks of the participation

No risk is foreseen. You are only requested to be available to participate.

Confidentiality and publication of the study data

Any responses you provide in the interview can be recorded or written down. The data, however, will not include any personal identification; hence it will not be possible to identify you afterwards. All the data you provide will be anonymised and treated confidentially. The information you provide will be analysed and presented in project reports together with the information from other participants. The raw data will be stored in the internal servers of the Know-Center protected by passwords that are only known to researchers conducting the interview. All the raw data will be stored for 5 years after the project finalisation.

Funding of the research

The research leading to this interview has received funding from the European Union's Horizon Europe Research and Innovation Programme, under Grant Agreement no 101120688.

Contact for further information or in case of withdrawal from the study

DI Dr. Angela Fessl, Know-Center GmbH, afessl@know-center.at

DRAFT

Consent Form

WP4 – T4.1 – Participant Information Sheet 03.2024

Consent Form

Background of this study

EMERALD is a Horizon Europe Project (GA no.: 101120688) with the objective to pave the road towards Certification-as-a-Service (CaaS) for continuous certification of harmonized cybersecurity schemas like the EUCS. This interview is conducted within WP4 – User Interaction and User Experience development of the EMERALD Project. The goal of this interview is to elicit requirements from our target groups such as auditors/chief information security managers/compliance managers etc. necessary for developing the integrated EMERALD UI. In more detail, our goal is to elicit in-depth insights about your work as auditors/chief information security managers/compliance managers in relation to continuous cloud auditing processes.

Statement of researcher's responsibility

As researcher, I have explained the nature of this research study and the procedures to be undertaken in this context. I have offered to answer any questions and fully answered such questions.

Declaration of participant

I confirm that: I am 18 years old or older and I am competent to provide consent. I have read and understood the information about this study, as provided in the Information Sheet. I have also had the opportunity to ask questions and all my questions have been answered to my satisfaction. I freely and voluntarily agree to participate in this research study. I understand that I may refuse to answer any question and that I may withdraw at any time without being penalised for withdrawing nor questioned on why I have withdrawn. I agree that my personal information will remain confidential and that my data will be used anonymously and securely in research and publications, in a way that my identity cannot be revealed. I understand that other researchers will have access to this data only if they agree to preserve the confidentiality of the data.
I agree to the terms and to the recording of the consent procedure/ and interview (phone interviews)

Participant:

_____	_____	_____
Name	Signature	Date

Researcher:

_____	_____	_____
Name	Signature	Date

© EMERALD Consortium Contract No. GA 101120688 Page 1 of 1
www.emerald-he.eu

CC BY-SA

Data Protection Information

Data Protection Information

Controller:	Know-Center GmbH Research Center for Data Driven Business & Big Data Analytics, Sandgasse 36/4, 8010 Graz Contact: info@know-center.at
Data protection officer:	Data Protection Officer of Know-Center GmbH Sandgasse 36/4, 8010 Graz Contact: datenschutz@know-center.at
Purpose of processing:	Maintaining business contacts to the extent that this is covered by the reasons for being contacted to which the data subject has consented.
Data:	Name, e-mail address, relevant for contacting the interview partners to which they have given their consent.
Basis in law:	Consent pursuant to GDPR Art 6 (1) (a)
Recipient:	No transmission to third parties; no contract processing
Transmission to third countries:	No
Duration of storage:	Until the time when you withdraw your consent. Irrespective of withdrawal of consent, the data will be deleted if your e-mail address becomes invalid or if we receive notification that communications are undeliverable.
Data subject rights:	You have the right to: - **Information and access**, to find out whether we have personal data of yours stored and what data it is. - **Rectification – correction and/or completion** of your personal data that are incorrect or incomplete - **Erasure – deletion** of your personal data that are being processed in a manner which is not lawful or is no longer lawful - **Restriction of processing** - **Data portability** - **Withdraw consent** that you have given, effective for the future: i.e. further processing of your data is then not allowed from that point in time onwards, unless there is an overriding legitimate reason for doing so. - **Object** to any assertion by Know-Center GmbH of an overriding legitimate interest in storing/processing the data To exercise these rights please contact datenschutz@know-center.at You also have a right to make a **complaint to the Data Protection Authority**. In this regard, we also refer to their homepage, which can be accessed under the link https://www.dsb.gv.at

DRAFT

CC BY-SA