



D8.11 – Data Management Plan

ORDP (Open Research Data Pilot)

V2.0

09/06/2022

Partner responsible for the document: ROTECH



Project Information

Grant Agreement Number	869529
Project Full Title	A Human-centred Internet of Things Platform for the Sustainable Digital Mine of the Future
Project Acronym	Dig_IT
Funding scheme	RIA - H2020-SC5-2019-2
Start date of the project	1 May 2020
Duration	48 months
Project Coordinator	María García Camprubí (ITAINNOVA)
Project Website	www.digit-h2020.eu

Deliverable Information

Deliverable n°	D8.11
Deliverable title	Data Management Plan
WP no.	8
WP Leader	CORE
Contributing Partners	All
Nature	ORDP (Open Research Data Pilot)
Authors	Marialuisa Scalise (ROTECH)
Contributors	Amal Alrish, Damiano Vallocchia (ROTECH) Isabel Marco, Ignacio Giménez (Ethics board)
Reviewers	WP Leader (CORE), Technical Manager (SUBTERRA) and Project Coordinator (ITAINNOVA)
Contractual Deadline	December 2020
Delivery date to EC	09/06/2022

Dissemination Level

PU	Public	X
PP	Restricted to other programme participants (incl. Commission Services)	
RE	Restricted to a group specified by the consortium (incl. Commission Services)	
CO	Confidential, only for the members of the consortium (incl. Commission Services)	

Document Log

Version	Date	Author	Description of Change
1.0	28.10.2020	Marialuisa Scalise	First issue ready for internal review
1.1	09.11.2020	Marialuisa Scalise	Implementation of suggestions proposed by CORE
1.2	02.12.2020	Marialuisa Scalise	Entry of partner contributions (datasets from IOTA, SINTEF, STRATAGEM, SUBTERRA, LIBRA, TITANIA)
1.2	09.12.2020	Konstantinos Nikolopoulos	First Review of DMP
1.3	18.03.2021	Noemi Tesauro, Marialuisa Scalise	Implementation of suggestion proposed by ITAINNOVA; entry of partner contributions (dataset from: SINTEF, SUBTERRA, ICCS, TAU)
1.4	13.04.2021	Diego Grimani, Marialuisa Scalise	Implementation of suggestion proposed by Project Coordinator
1.5	14.04.2021	Konstantinos Nikolopoulos, Andreas Mantelos, Ioannis Livieris	Second Review of DMP
1.6	18.03.2022	All partners	Intermediate issue ready for internal review. The following paragraphs have been added: Data Lifecycle in Subsection 2.2; FAIR Data Principles in Subsection 2.3 and Resources, Security and Ethics in Section 4. With the contribution of all partners and according to PO remarks, the following information have been inserted/updated: Data collection, access and sharing procedures, how each partner is handling the data to make it FAIR, in Section 3.1 and datasets in Subsection 3.2.
1.7	22.03.2022	CORE	Review as WP Leader
1.8	19.05.2022	ROTECH, Ethics board	Implementation of project ethics board remarks
2.0	09.06.2022	ROTECH	Minor changes.

Table of contents

1	Introduction	8
1.1	Scope	8
1.2	Project Overview	8
1.3	Document organisation	9
2	Guidelines	10
2.1	DMP general definition	10
2.2	Data lifecycle	10
2.3	FAIR data principles	11
2.4	Guidelines for the dataset description	13
3	Data Management Plan	15
3.1	General	16
3.1.1	Handling of research data during and after the end of the project	18
3.1.2	What data will be collected, processed and/or generated	20
3.1.3	The methodology and standards applied	22
3.1.4	Sharing the data	24
3.1.5	Preservation of the data during and after the end of the project	25
3.2	Dig_IT Datasets	27
3.2.1	Dig_IT Use Cases datasets	27
3.2.1.1	UC1 – Marini Marmi (underground quarry)	28
3.2.1.2	UC2 – La Parrilla mine (open pit)	37
3.2.1.3	UC3 – Titania AS (open pit)	39
3.2.1.4	UC4 – Hannukainen mine (open pit)	49
3.2.1.5	UC5 – Sotkamo Silver mine (underground mine)	52
3.2.2	Other datasets	61
4	Resources, Security and Ethics	68
4.1	Allocation of resources	68
4.2	Data security	69
4.2.1	Microsoft Teams	69
4.2.2	Project Website	70
4.3	Ethical aspects	71
5	Opinions from the ethics board regarding this deliverable	75
6	Conclusion	76
7	References	77

List of Figures

Figure 1 Data Lifecycle stages	11
--------------------------------	----

List of Tables

Table 1 Dataset table template.....	13
Table 2 Summary of the significant research data categories	15
Table 3 Handling the research data during and after the end of the project.....	18
Table 4 The data which will be collected, processed and/ or generated.....	20
Table 5 The methodology and the standards applied in the project	22
Table 6 Sharing the data in the project.....	24
Table 7 Preservation of the data during and after the end of the project.....	25
Table 8 Datasets for each use case	27
Table 9 Marini (UC 1) – Geotechnical Data – Geotechnical Dataset for Digital Twin.....	28
Table 10 Marini (UC 1) – EHS & Sustainability data - OHSE Dataset	29
Table 11 Marini (UC 1) – Situational awareness & biometrical data - Stress & Fatigue Monitoring Dataset .	30
Table 12 Marini (UC 1) – Situational awareness & biometrical data - Location Dataset	31
Table 13 Marini (UC 1) – Situational awareness & biometrical data - Wearable OHSE Dataset	32
Table 14 Marini (UC 1) – Digital Twin Air Quality Data	33
Table 15 Marini (UC 1) – Situational awareness & biometrical data - Smart Garment Command & Control Dataset.....	34
Table 16 Marini (UC 1) – Assets data	35
Table 17 Marini (UC 1) – PMA Data.....	35
Table 18 Marini (UC 1) – POS Data.....	36
Table 19 Marini (UC 1) – Intelligent Toolbox for OHSE Data.....	37
Table 20 La Parrilla (UC 2) – Geotechnical Data – Geotechnical Dataset for Digital Twin	38
Table 21 Titania (UC 3) – Geotechnical Data – Geotechnical Dataset for Digital Twin.....	40
Table 22 Titania (UC 3) – EHS & Sustainability data - OHSE Dataset	42
Table 23 Titania (UC 3) – Situational awareness & biometrical data - Location Dataset	43
Table 24 Titania (UC 3) – Situational awareness & biometrical data - Wearable OHSE Dataset	44
Table 25 Titania (UC 3) – Digital Twin Air Quality Data	45
Table 26 Titania (UC 3) – Situational awareness & biometrical data - Smart Garment Command & Control Dataset.....	45
Table 27 Titania (UC 3) – Assets data - Trucks Dataset.....	46
Table 28 Titania (UC 3) – PMA Data.....	47
Table 29 Titania (UC 3) – POS Data	48
Table 30 Titania (UC 3) – Intelligent Toolbox for OHSE Data.....	48

Table 31 Titania (UC 3) – Intelligent Toolbox for OHSE Data.....	49
Table 32 Hannukainen (UC 4) – EHS & Sustainability data - Water quality, laboratory results dataset	50
Table 33 Hannukainen (UC 4) – EHS & Sustainability data - Water quality, online measurement dataset ...	50
Table 34 Hannukainen (UC 4) – EHS & Sustainability data - Weather station dataset	51
Table 35 Sotkamo (UC 5) – EHS & Sustainability data - OHSE Dataset.....	52
Table 36 Sotkamo (UC 5) – Situational awareness & biometrical data - Stress & Fatigue Monitoring Dataset	53
Table 37 Sotkamo (UC 5) – Situational awareness & biometrical data - Location Dataset.....	54
Table 38 Sotkamo (UC 5) – Situational awareness & biometrical data - Wearable OHSE Dataset.....	55
Table 39 Sotkamo (UC 5) – Digital Twin Air Quality Data.....	56
Table 40 Sotkamo (UC 5) – Situational awareness & biometrical data - Smart Garment Command & Control Dataset.....	57
Table 41 Sotkamo (UC 5) – Assets data – Maintenance log Dataset.....	58
Table 42 Sotkamo (UC 5) – PMA Data	58
Table 43 Sotkamo (UC 5) – POS Data.....	59
Table 44 Sotkamo (UC 5) – Intelligent Toolbox for OHSE Data	60
Table 45 Sotkamo (UC 5) – Intelligent Toolbox for OHSE Data	60
Table 46 Data for the Life Cycle Assessment (LCA).....	61
Table 47 Public website data.....	62
Table 48 Aggregator dataset coming from assets and going to ICCS platform	63
Table 49 Alerting-related Data.....	64
Table 50 Environmental emission data written in the Blockchain	66
Table 51 End-User Needs dataset description.....	66

Acronyms table

Acronym	Description
API	Application Programming Interface
CA	Consortium Agreement
DMP	Data Management Plan
GDPR	General Data Protection Regulation
EHS	Environment, Health & Safety
FAIR	Findable, Accessible, Interoperable, Re-usable
H2020	Horizon 2020
IIoT	Industrial Internet of Things platform
IPR	Intellectual Property Rights
OHSE	Occupational Health, Safety and Environment
ORDP	Open Research Data Pilot
UC	Use Case
UDA	Unscheduled Downtime of Assets
PDM	Predictive Maintenance Agents
POS	Predictive Operation System

1 Introduction

1.1 Scope

A Data Management Plan (DMP) is a formal plan that outlines what data will be generated or collected, how data will be managed (access, storage, backup), the standards in use, the workflow to make the data accessible for use, reuse and verification and which plans for data sharing and preservation exist ensuring that data are well-managed. In addition, the purpose of this DMP is to help demonstrators to manage the data and help other potential users to use the shared data.

The initial version of the DMP submitted in M8 lacked some data due to covid pandemic which has a great impact on the WP2 tasks related to data generation and gathering. Thus, this deliverable is an updated version of the initial DMP, which is in line with the H2020 guidelines for data management plan creation and identifies the initial classes of datasets of the project. This updated version reflects the updates on data management procedures during the first 24 months of the project.

This DMP covers the whole research data life cycle. A knowledge management system will be created, based on DMP, describing how acquired data and knowledge will be shared and/or made open, and how it will be maintained/preserved.

Identifiable project data will be provided in a manner to define relevant knowledge, increase partners' awareness, validate the result, and timeframe of actions. All partners are responsible for periodically updating the knowledge management system with outcomes of research work conducted by other groups and any new patents, i.e., to ensure on-going relevant technological developments and innovations are identified, analysed, and built upon during the course of the project.

1.2 Project Overview

Dig_IT – A human-centred Internet of Things platform for the sustainable digital mine of the future – is a Horizon 2020 funded project that will address the needs of the mining industry to move forward towards a sustainable use of resources while keeping people and environment at the forefront of their priorities. In order to achieve that, Dig_IT proposes the development of a smart Industrial Internet of Things platform (IIoTp) that will improve the efficiency and sustainability of mining operations by connecting cyber and physical systems. The platform will collect data from sensors at 3 levels: human (which are only "biometric data processed solely to identify a human being"), assets, environment and will also incorporate both market real time and historical data.

The Dig_IT project aims at contributing to the introduction of innovative concepts and technologies towards the vision of the European Digital mine of the future. Hence, the achievements are required to be communicated to the public and the respective targeted markets. The results are required to be distributed openly to the public and as a result Dig_IT will participate in the Open Research Data Pilot in Horizon 2020. Dig_IT partners are aligned with EU targets to improve and maximise access and reuse of research data generated by Horizon 2020 projects and work proactively to ensure data's openness, safety and commercialization.

However, the quality and performance data of the consortium's enterprises will be considered private and will only be available after granting permission.

This being said, where possible and where there are no particular ethics board concerns (biometric and personal data from workers, for instance), research data generated and collected during the course of the project will be freely dully and suitably anonymised or pseudonymised distributed by means of, for example, technical journals which requires publishing data to accompany the article.

1.3 Document organisation

The document is organised in the following four sections, according to the template provided by the European Commission [1]:

- 1- Introduction: This section gives general information about the project and the document itself.
- 2- Guidelines: In this section there is a general description of DMPs in the concept of an H2020 project and a short summary of the guidelines that have been provided to the involved partners and end-users for the provision of information on the datasets of which they are producers or users. In this section, it has also been inserted a typical data lifecycle description and a paragraph to explain the FAIR data principles.
- 3- Data Management Plan: This is the core section of the document. Data collection, access, and sharing procedures will be presented in this document together with the data management plan for distributed management technology. Moreover, each partner has described how they are handling the data in order to make the research data FAIR. Furthermore, it contains the datasets specification organised per Use Cases (UC), other datasets and a sub paragraph to describe Common datasets.
- 4- Resources, Security and Ethics: This section contains the resources allocated for data management and the data security and ethical aspects considered in the project.
- 5- Conclusions: This section describes the current status of the document.

2 Guidelines

2.1 DMP general definition

Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR, see below for details), a DMP should include information on:

- the handling of research data during and after the end of the project (refer to section 3.1.1),
- what data will be collected, processed and/or generated (refer to section 3.1.2),
- which methodology and standards will be applied (refer to section 3.1.3),
- whether data will be shared/made open access (refer to section 3.1.4),
- how data will be curated and preserved, including after the end of the project (refer to section 3.1.5).

2.2 Data lifecycle

Characteristics and features of all the data generated, collected, processed, analysed, stored, and/or eliminated during any of the data lifecycle stages must be defined. Herein, a workflow to be followed to make data FAIR (Findable, Accessible, Interoperable and Reusable) and specifications about data preservation (or elimination) are defined.

The management of research data is a fundamental part of any research process in order to provide a clear overview of the data life cycle and to facilitate accessing and reusing data, not only during the project, but also in the future.

The data lifecycle of a project includes six phases: creation of data, data storage (repositories), use of data, sharing (between users, partners, or external researchers), archive (whenever data are stored for future activities) and destruction of data (if necessary).



Figure 1 Data Lifecycle stages

In our project, we address data management during every lifecycle stage: (1) from the collecting data point, (2) the generation, modification, or cleaning of data during its processing (project common platforms), and (3) the storing relevant data and metadata to promote sharing and reusing data for the aim of collective intelligence, including (4) keeping data transactions in the distributed ledger technology.

As far as it concerns use cases, different data types must be dealt with depending on the individual requirements of each use case. This fact makes it necessary for different processing and storing tools to be available and defined, as well as what file formats to be used.

If necessary, data management should go deeper in this context and also define other data characteristics, i.e., container architecture, usage of standard file formats, data loss risks, among others, as well as to provide understanding about the potential risk of using the proposed blockchain technology to ensure that full comprehension and management of data are accomplished during the project.

2.3 FAIR data principles

This document should ensure that the partners follow the “FAIR Guiding Principles for scientific data management and stewardship” [2]. These principles set up the standards for data to be easily findable, accessible, interoperable and reusable not only by humans, but also for computational systems.

These FAIR Guiding Principles are presented below, as they were published in the original source:

To be Findable

- F1. (meta) data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta) data are registered or indexed in a searchable resource

To be Accessible

- A1. (meta)data are retrievable by their identifier using a standardized communication protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable

- I1. (meta) data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (meta) data use vocabularies that follow FAIR principles
- I3. (meta) data include qualified references to other (meta)data

To be reusable

- R1. (meta) data are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta) data are released with a clear and accessible data usage license
 - R1.2. (meta) data are associated with detailed provenance
 - R1.3. (meta) data meet domain-relevant community standards

These FAIR principles provide a definition of the practices and methods that data resources and infrastructures should follow in order to support discovery and research.

Upon completion of each dataset, the following questions should have been answered to address FAIR data principles:

Findability: 'Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism?', and 'What metadata will be created?'

Accessibility: 'Which data produced and/or used in the project will be made openly available as the default?', and 'How will the data be made accessible?'

Interoperability: 'Are the data produced in the project interoperable?'

Re-usability: 'How will the data be licensed to permit the widest re-use possible?'

2.4 Guidelines for the dataset description

Table 1 represents the template that the partners have to use to describe each individual foreseen dataset. Moreover, at the right of item to be filled in this template, we introduced some explanation information about the contents required to facilitate its filling. This table (empty) has been delivered to each partner and the contribution of each partner in charge of handling data is in chapter 3 and will represent the core content of the DMP.

Table 1 Dataset table template

Dataset reference and name	Unique identifier and name for the produced data to grant the discoverability.
Description	Insert a description of the content of the dataset, origin, nature, scale and, if possible, purpose. Enter a reference to a potential audience of interest.
Source	Meeting, survey, indicators, sensors, etc.
Other partners contributing	Indicate the Partners involved or N/A.
Personal data	This is a binary field (Yes/No) that indicates whether the dataset contains personal data or not.
License	Insert terms and conditions under which the dataset is provided, if applicable.
Embargo periods	Indicate the periods or N/A.
Legal framework(s)	Indicate (if any) the legal framework(s) regulating or affecting in any way this dataset. And how is this addressed or N/A.
Technology	Indicate technology adopted to expose the data asset / API (preferring non-proprietary formats accessible via open-source software).
Standards and Metadata	Describe the data that will be created by the research project, in terms of content and quality. Describe the procedures used to create the data collection (e.g., comparison between data sources, creation from scratch, analysis of existing cases, etc.).

Dataset reference and name	Unique identifier and name for the produced data to grant the discoverability.
	Describe the metadata that will accompany the dataset and specify the standards used for the metadata (good documentation increases the detectability of data) ¹ . Provide sufficient information aimed at the self-description of the dataset, thus offering an effective interpretation of the context of its production (e.g., summary document attached to the dataset with examples, description of the research focus, dates and references of origin).
Data sharing	State the conditions under which data are shared, including how users will be able to access these data (access procedures), dissemination strategies, range of open data, embargo periods.
Data volume	Define expected volume of dataset.
Data type and format	Define the type and format of the data, justifying, if possible, the choice (e.g., conversion of data useful for conservation purposes, use of standard formats accessible through open-source software).
Update frequency	Indicate the period of time which the data is updated (if foreseen).
Archiving and presentation	Describe how data will be presented and archived, and, if possible, why such option has been chosen, their storage and backup, long-term preservation and ways to achieve it, approximated end volume and plans to be covered.
Responsibilities	Define the team responsible for data management and data curation.
Resources	Define the resources necessary for data acquisition, elaboration and maintenance. Consider that the definition of resources varies proportionally to the volume of data, therefore to the quality and format to be maintained, as well as to the storage and backup systems that you want to prepare.
Additional services and provisions	Services for database hosting, data integration, data visualisation, data security and so on.
Ethical aspects	Indicate whether there are any ethical or legal issues that can have an impact on data sharing such as whether consent for data preservation is gained or not, whether the data is protected by database law or copyright, personal data and so on.
Other	Refer to other national/funder/sectorial/departmental procedures for data management that are used on the dataset (if any).

¹Metadata could be defined as the information associated with the data to identify them and describe their context, content and structure, as well as to allow their management over time in the storage system.

3 Data Management Plan

During the proposal phase a preliminary DMP was appointed.

Table 2 shows a summary of the significant research data categories identified so far that will be collected and generated during and after the Dig_IT project life cycle.

Table 2 Summary of the significant research data categories

Research Data	Leading partners	Collaborating Partners
Geotechnical data of the mine sites (e.g., orientation, spacing etc.)	SUBTERRA	Titania AS, La Parrilla and Marini Marmi, BRUNEL
EHS (e.g., dust and noise), Sustainability (e.g., water analyses)	SINTEF	Titania AS, Marini Marmi Srl, Tapo, Schneider Electric, EUROCORE, BRUNEL
Smart garment user localization, Smart garment user biometrical data (heart rate, skin temperature, etc.), Smart garment wearable OHSE measurements (temperature, humidity, gas), Smart garment headset subsystem (keyword detection, noise)	ICCS	Titania AS, Marini Marmi Srl, TAPO (Sotkamo), CORE-INNOVATION, LIBRA, IOTA, EUROCORE
Asset Data available data from use cases related to: Extraction process, Mineral processing & treatments, Energy consumption, Water consumption, Asset data, EHS measurements, OHSE monitoring etc.	TAU, SUBTERRA, CORE	Titania AS, Tapo, Schneider Electric, BRUNEL
Data for the Life Cycle Assessment (LCA)	STRATEGEM	SINTEF, Marini Marmi Srl, Tapo
Early warning hazard prediction datasets.	BRUNEL	Titania, Marini Marmi Srl
Evolutionary Algorithm datasets.	BRUNEL	Titania, Marini Marmi Srl
Data aggregation point from sensors and mobile assets	SCHNEIDER ELECTRIC	
Alerting – related data. For each triggered alert, the collected data are the timestamps, the metric that the alert concerns and the value of the metric that triggered the alert. All the alerts are defined based on data available in the data warehouse.	LIBRA	ICCS
Data generated by the virtual sensors (Air and water quality)	ITAINNOVA	SINTEF, SCHNEIDER, LIBRA, MARINI MARMINI, SOTKAMO, TITANIA
End user needs	CORE	TAPOJÄRVI OY, Titania AS, Subterra, Marini

Research Data	Leading partners	Collaborating Partners
Predictive Operation System (POS) predictions	CORE	Marini Marmi Srl, Titania AS, TAPO (Sotkamo)
Predictive Maintenance Agents (PMA) predictions	CORE	Marini Marmi Srl, Titania AS, TAPO (Sotkamo)
Intelligent Toolbox for OHSE	LIBRA	Marini Marmi Srl, Titania AS, TAPO (Sotkamo), CORE

3.1 General

This Data Management Plan (DMP) has been prepared by taking into account the template of the “Guidelines on Data Management in Horizon 2020”. Elaboration of the DMP will allow Dig_IT partners to address all issues related to management of data collected during the project as well as ethics. DMP was planned as a deliverable for M8. However, the document has been updated up to M24 to collect a more complete description of the project datasets.

The consortium will comply with the requirements of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation) on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Type of data, storage, recruitment process, confidentiality, ownership, management of intellectual property and access: The Grant Agreement and the Consortium Agreement are to be referred to for these aspects. The procedures that will be implemented for data collection, storage, access, sharing policies, protection, retention and destruction will be according to the requirements of the national legislation of each partner and in line with the EU standards.

The Advisory Board of the project will also ensure that EU standards are followed. Informed consent will be provided to all participants in the project data collection processes. All collection of sensitive data will be done with full consideration of data protection principles and will satisfy data protection requirements in accordance with EU and non-EU directives and national implementations thereof. Due to nature of services, it is NOT likely that personal will be captured and processed. In case that there will be sensitive/ personal data, collection and processing will be done according to the applicable data protection provisions, such as Regulation (EU) 2016/679 on the protection of individuals with regard to the processing of personal data and on the free movement of such data including article 29 working group 8/2010 opinion and Directive 2002/58 on Privacy and Electronic Communications.

For this reason, in case of personal data collection and processing, anonymous and pseudonymous user data will be collected and securely stored. Anonymous identification of user-provided

information will be leveraged only to confirm the authenticity of users interacting with the system and to prevent malicious behaviour. In addition, raw data coming from sensors are not stored directly to a database, but rather feed algorithms that are used to identify events. These events are anonymized in such a way so as not to be linked to certain individuals, thus will not fall within the scope of the GDPR. Pseudonymous user data, on the other hand, are needed in order to train and validate algorithms related to workers' fatigue and stress levels (See D10.2 for further information).

No need to personally identify users through their information is envisioned nor to include sensitive data. The collected data will be treated anonymously and additionally a various set of measures will be put in place in order to protect user privacy and its data security, by embedding privacy by design principles from the early stage of the project technical start. Where needed, a prompt Privacy Impact Assessment (PIA) exercise will be performed.

The type, quality and quantity of accessed data will be regulated, by designing and implementing adequate PIR (Privacy Information Retrieval) and PPQ (Privacy Preserving Query) mechanisms.

By referring to the proposed work plan, it is worth noticing that all such measures will be considered at all levels of the technical project development, starting from WP1 (Project specification and industry needs), from WP2 where data gathering will be done, up to WP3, WP4 and WP5 where data will be used to build different algorithms and services. While new required and relevant technologies will be developed as part of the project, Dig_IT is already aware of the following existing technological measures necessary to minimise associated privacy risks such as:

1. Use of secure data storage, encrypted transfer of data over the capturing channels, controlled and auditable access for different classes of data;
2. Obscuring/removing user identities at the source of field trial data generation to prevent direct user tracing;
3. Obscuring personal location data through indirect or delayed routing to prevent individual localization as much as possible and limit user tracking through correlation of depersonalised data based on its location.

DMP will define how data will be shared and more specifically the access procedures, the embargo periods, the necessary software and other tools for enabling re-use, for all datasets that will be collected, generated, or processed in the project. In case the dataset cannot be shared, the reasons for this will be mentioned (e.g., ethical rules of personal data, intellectual property, commercial, privacy-related, security-related). In addition, beneficiaries do not have to ensure open access to

specific parts of the research data if the achievement of the action's main objective, would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management will present the reasons for not giving access.

As we said previously, in order to make research data FAIR, each partner described the aforementioned 5 points the DMP should include in Table 3, Table 4, Table 5, Table 6, Table 7 in the following subsections respectively.

3.1.1 Handling of research data during and after the end of the project

Table 3 Handling the research data during and after the end of the project

Partner name	Handling the research data during and after the end of the project
SUBTERRA	The data management and access it will just take part of employees of SUBTERRA involved in the project only with the private server of SUBTERRA. After the project, as we take part of the Technical Management, we will still have access to the real time data with the same methodology.
SINTEF	During and after the project some data will be stored on our own servers for the purpose of calibration and evaluation of the data sampling. In the initial stage some data will be stored on an external service from Mathworks. This is for testing only and no data will be traceable to the project or its partners. Before the sensors are applied to the end user, the data management is transferred to the Dig_IT-platform and no longer managed by SINTEF. Some of the sensor bundles can save data locally to a SD-card. If needed and approved by the end user, this data may be sent to SINTEF for evaluation and stored on our own servers.
TAU	During the project, the data collected from Tapojärvi mines will be stored to a server located at the Tampere university. Data is available to named Dig_IT researchers at the university. In addition to that, data is streamed to Schneider aggregator. After the project, the server and the data is transferred to the Tapojärvi mine.
CORE	Some data will be stored on our own servers for the duration required by Dig_IT project in order to develop and fine-tune the AI-based prediction models. This data will not be traceable to the partners of the project. User needs data are stored in project's TEAMS repository.
ITANNOVA	All the data will be directly generated into the Dig_IT platform. No data will be collected during nor after the project duration by ITAINNOVA.
ICCS	During the timespan of the project the data accumulated from the smart garment and other Dig_IT sensors will be stored in the ICCS platform (Dig_IT-platform), i.e., the data warehouse. It will include data that require long term storage, but also real time data with no storage requirements, i.e., send and delete. Dig_IT partners will have access to raw (and processed) data that will be exploited for further processing and visualisation to provide insights for the use cases that have been identified per mine. This process will facilitate the

Partner name	Handling the research data during and after the end of the project
	realization of the trials and evaluation of technologies exploited in the project, and thus their potential to enhance operations in the mining industry. At the end of the project, the data handling (storage, deletion. etc.) will be decided by the respective mines and relevant project partners.
LIBRA	All the data used or produced by LIBRA are directly consumed from the ICCS platform, i.e., the data warehouse. The purpose is just to visualise aspects of these data and they are not get stored in any hard disk or similar media connected to the decision support system. On the other hand, LIBRA generates data related to the triggered alerts which are stored to the ICCS platform following the corresponding data handling strategies.
STRATEGEM	Only during the project STRATAGEM will handle research data. The reason for handling this data is to provide the mine owners with an outcome of the environmental life cycle perspective after the installation of the Dig_IT technologies. Therefore, there is no need for STRATAGEM to handle this data after accomplishing this purpose. STRATAGEM will handle data after the end of the project only in the case that a mine would like to see supplementary Life Cycle Assessment results.
Titania	Titania is an end-user in this project and as such is a data owner and -provider to the project. Titania will handle its own data during the project, and handle the resulting research data related to Titania after the project according to internal data management at Titania.
Marini	Marini as end-user in the project is a data owner and provider to the project. Marini will handle its own data during the project and the results of the data related to Marini use case after the project.
TAPOJÄRVI	Tapojärvi is an end-user in this project and as such is a data owner and provider to the project. Tapojärvi will handle its own data during the project, and handle the resulting research data related to Tapojärvi after the project according to internal data management at Tapojärvi.
ROTECH	Ro Technology is responsible for creation and managing the project website. Thus, ROTECH receives the outcomes and up-to-date information about the project from the partners, organize and publish them on the website. The website will be available during the project period and one year after the end of the project.
BRUNEL	Data will be securely stored on Hard Disks during the project period. Data may be transferred using secure standard techniques (sftp, USB, etc) among members of the project. Handling/managing the data after the end of the project will be decided in cooperation with the other partners.
SEI	Only during the project, SCHNEIDER ELECTRIC will handle research data. The reason for handling this data is to test the communication interface before the installation of the Dig_IT technologies. The data (that not included personal data) will be managed only during the project and only from researcher involved in the project. At the end of the project the data will be managed only from the end-user.

Partner name	Handling the research data during and after the end of the project
EUROCORE	All the data to which EUROCORE may have access, will be produced by ICCS, SINTEF. The purpose is to support the reviewing / optimisation of OSH related alerts and give more accurate input in the intelligent toolbox for OSHE. Therefore, there is no need for EUROCORE to handle this data after accomplishing this purpose.
IOTA	Dig_IT will use blockchain permanodes for long-term data storage by implementing filters to store only specific information. The data in permanodes will be encrypted and accessed only by the data owner. The network currently deployed is permissionless, IOTA nodes can be run also by parties other than those part of the Dig_IT environment and permanodes can be hosted by everyone in the world. However, data related to different data sources in Dig_IT will be accessible only within specific encrypted channels, which are known and accessible only by entities who have been verified as party to the Dig_IT environment and rightful entity for data processing. IOTA will provide this infrastructure to store the hashed data and encrypted non personal dataset during the project duration that can span after the project lifetime in case this is required by the partners. The data are not personal, and it will be accessed by the project partners, with a limited subset of these data available online for public and auditors' transparency.

3.1.2 What data will be collected, processed and/or generated

Table 4 The data which will be collected, processed and/ or generated

Partner name	The data which will be collected, processed and/or generated
SUBTERRA	Will collect data to assess geotechnical risks. In situ data acquisition will be request, also we will get Multispectral and RGB imagery from UAV flights as well as other geotechnical sensors data (piezometers, pressure cells, radar ...). To start monitoring all these information, we will have to be able to get this data in real time. For the moment we have just analyse of the historical and previous data.
SINTEF	The sensor bundles made by SINTEF will be equipped to measure OHSE parameters, as particulate matter (PM), gases (CO ₂ , NO _x , etc.) , temperature, humidity, and possibly noise. For the calculation of metals or a particular mineral fraction, calibration parameters specific to each site must be applied. SINTEF may when approved by the end user ask for and use data collected by the Dig_IT-platform to apply and evaluate calibration factors. The data from the sensors will be managed by the Dig_IT-platform and not by SINTEF. Data provided to SINTEF will be stored on SINTEFs own servers.
TAU	We are collecting the data straight from the vehicle's own system. Data includes i.e. engine speed, fuel rate and GPS. The gateway that collects the data and sends it to the edge server, is installed to the target vehicle by us. Any GDPR data is not collected. Also, any additional sensor, besides the vehicle's own sensors, are not installed by TAU.
CORE	Data store in the Dig_IT IIoT platform Data Warehouse (assets' data, data from the smart garment & air quality sensors' data) will be collected and

Partner name	The data which will be collected, processed and/or generated
	processed for the development of the AI-based models. The generated outputs of these models will concern predictions about the health state, energy consumption, downtime cost and maintenance cost of a identified asset, health state of an individual as well as future gases and PMs concentrations. Additionally, CORE collected data from the end users to identify their market needs.
ITANNOVA	The data will be generated by the virtual sensors, which will be created based on computational fluid dynamic simulations (CFD). These virtual sensors will be embedded into the DIG_IT platform where they will be evaluated using data coming from physical sensors, and will provide data of water and air quality. ITAINNOVA will not collect nor treat any of this data.
ICCS	The smart garment is equipped with components to measure various OHSE parameters. It will be equipped with sensors that will monitor the environmental variables, such as air quality (CO, CO2, NOx, NH3 gases), temperature, relative humidity and noise level. A localization module using ultra-wide band (uwb) technology will measure the position of the user within the mines. The smart garment will also incorporate a smart wristband subsystem which will measure biosignals such as photoplethysmogram (PPG), skin temperature, electrodermal activity (perspiration) as well as motion (accelerometer and gyroscope). Lastly, the smart headset subsystem will be capable of detecting certain keywords using voice recognition. Please follow D2.4 for more details on the smart garment sensors.
LIBRA	LIBRA's Decision support system distributes alerts based on the consumed data from the ICCS platform. When an alert is triggered, the corresponding alert id, the timestamp of the triggering time, as well as the metric value responsible of triggering the events are generated and stored back in the ICCS platform. Moreover, LIBRA processes the data collected in the ICCS platform for data visualization purposes.
STRATEGEM	Datasets will be collected for the task 5.2 LCA modelling procedure. For the needs of this task, STRATAGEM will first collect average historical data of the mine operation in all the life cycle stages. Those data will be compared against the collected data after the installation of the IIoTp platform. The generated data/results will show the environmental impact before and after the installation of IIoTp platform to see at the end if the new installations affected the mine operation in practice in terms of environmental impact. The second part of this study includes the collection of all the energy and resources consumed for installing the IIoTp platform in the mines. This dataset will assess the Life Cycle environmental impact of the platform and those will be the generated results.
Titania	Titania is an end-user in this project and will provide data according to table 2 and the request of the task leading partners.
Marini Marmi	Marini is an end-user in the project and will provide data according to table 2 and to the requests of other partners. Data from Marini will be provided to other partners.

Partner name	The data which will be collected, processed and/or generated
TAPOJÄRVI	Tapojärvi is an end-user in this project and will provide data according to table 2 and the request of the task leading partners.
ROTECH	The data are up-to-date information about the project, relevant outcomes from the meeting, Milestones, project documents, and partners' press release.
BRUNEL	Mainly, data obtained from the relevant use-case partners (Titania and Marini) will be processed. This includes the Geotechnical data from SUBTERRA, EHS & Sustainability from SINTEF and Assets data from TAU. Some data will be generated by applying algorithms to the data obtained from these partners.
SEI	The data comes from mobile assets and sensor (trucks, drone , sensors). They will be stored in a platform and share with ICCS platform by MQTT and stored in SQL database.
EUROCORE	The data to which EUROCORE may have access (not collection, production or treatment) concern the OSH related measurements taken from the smart garment (ICCS) and the sensors used by SINTEF.
IOTA	All raw data generated during the project collected from the IOT devices, as well as data after pre-processing (calibration), excluding the personal data from the smart garment. The generated data/results will ensure the immutability, after they produced hashes of the data are written into the blockchain.

3.1.3 The methodology and standards applied

Table 5 The methodology and the standards applied in the project

Partner name	The methodology and standards applied
SUBTERRA	We are applying modelling technology with Leapfrog and Flac3D where we will be able to make the digital twin with all the data, we are getting from the UAV flights and other geotechnical historical data. After that in the model we will add all the sensors data and UAV flights to make the real time digital twin.
SINTEF	The sensor bundles are made with low-cost electronics and built around the Arduino or equivalent microcontroller. The protocols used for data collection in the unit is mainly I2C, a synchronous, multi-controller/multi-target, packet switched, single-ended, serial communication bus invented in 1982 by Philips Semiconductors. It is widely used for attaching lower-speed peripheral ICs to processors and microcontrollers in short-distance, intra-board communication. For sending data to the Dig_IT-platform MQTT, an IoT messaging protocol is used. This is done via an internet connection using WiFi or Ethernet.
TAU	The data is collected from the vehicle by using J1939 and CANopen protocols. Data is transferred as JSON format to the edge server by using HTTPS protocol. OAuth2 is used for authorization. Time synchronization between the gateway and the edge server is implemented by NTP.

Partner name	The methodology and standards applied
CORE	<p>The generated data will be composed by the predictions of the developed AI-based models. The models will be based on state-of-the-art neural network architectures (Encoder-Decoder, LSTM, Transformer) for the prediction of identified KPIs. The generated data will be published to Dig_IT-platform using MQTT/Kafka pub/sub messaging model, which constitutes an IoT well defined protocol.</p> <p>End-User Needs data were collected online office form. Questions were based on market research, market trends and project's innovations characteristics.</p>
ITANNOVA	<p>We are using an own methodology which is based on the combination of numerical simulations (CFD) and reduced order model techniques (ROM). The input variables driving the case of study are considered to build a design of experiments (DOE) which is evaluated by a computational fluid dynamic software (OpenFoam).</p> <p>The data generated with the CFD models is used to build the digital twin using an in house ITAINNOVA ROM library (Twinkle).</p> <p>Data coming from physical sensors located at few discrete locations is used to evaluate the digital twin, which is capable to generate data at any point of the considered case of study.</p>
ICCS	<p>Data reaching the ICCS platform will be exploiting protocol MQTT: The Standard for IoT Messaging. It is designed as an extremely lightweight publish/subscribe messaging transport that is ideal for connecting remote devices with a small code footprint and minimal network bandwidth. MQTT today is used in a wide variety of industries, such as automotive, manufacturing, telecommunications, oil and gas, etc. Additionally, partners will be granted access to data aggregated at the ICCS platform though a dedicated PostgreSQL TimescaleDB database to efficiently handle data requiring (relatively) long storage period, as well as access to DIG_IT Kafka streams to exploit (near-)real time data from the project's sensors.</p>
LIBRA	<p>We only consume data directly from the ICCS platform in two ways. Either through the PostgreSQL TimescaleDB or through the Kafka broker.</p>
STRATEGEM	<p>Methodology -> We are following the ISO 14040, which describes the principles and framework for LCA, and the ISO 14044 which specifies requirements and provides guidelines for LCA. Both standards focus mainly on the process of performing LCA, following a product's impact from cradle to grave. Additionally, the International Reference Life Cycle Data System (ILCD) handbook, which was developed by the European Commission Joint Research Centre, was followed for guidance, as it offers technical guidelines on conducting detailed LCA studies and it is consistent with the ISO 14040 and 14044 LCA standards. Technology ->The collected data are imported into the Software SimaPro version 9.2, which is a widely used commercial LCA software. SimaPro is used for the LCI and LCIA procedures together with the LCI database Ecoinvent version 3.</p>
Titania	<p>Titania is an end-user in this project and as such is a data owner and provider to the project. This question is not applicable for Titania.</p>

Partner name	The methodology and standards applied
Marini Marmi	Since Marini Marmi is just an end-user in the project, this question is not applicable for Marini Marmi.
TAPOJÄRVI	Tapojärvi is an end-user in this project and as such is a data owner and provider to the project. This question is not applicable for Tapojärvi.
ROTECH	Ro Technology uses WordPress (last release), a free open-source content management system for websites development, to create and manage the project website.
BRUNEL	We are applying Big Data Optimization for early warning hazard prediction and Phase-based Evolutionary Computation for smart scheduling tools. Mainly AI/ML techniques will be applied to data.
SEI	We are using Ecostruxure Machine Scada Expert with MQTT drivers and a SQL Database. The platform run in a Schneider industrial PC with Windows OS.
EUROCORE	Not applicable
IOTA	IOTA is using W3C decentralized identities for identification of the devices that generates the data, as well as to identify other system components, such as IIOT gateway. The IOTA DLT is also compliant with ISO/TC 307 namely ISO/AWI 7603 Decentralized Identity standard for the identification of objects by providing SSI based on decentralized identities.

3.1.4 Sharing the data

Table 6 Sharing the data in the project.

Partner name	Data Sharing
SUBTERRA	We receive the data by e-mail, where we will manage the data in our private server and share the information by the Teams Dig_IT server to share all the data with the other members of the project.
SINTEF	SINTEF will not share any data without written consent from the end users.
TAU	We will share the data with Schneider aggregator primarily with HTTPS using Kafka. Kafka handles the data sharing and makes sure that the data is as real-time as possible.
CORE	The generated data will be composed by the predictions of the developed AI-based models. This data will be shared within Dig_IT IIoT platform under platform's constraints, authorization and security.
ITANNOVA	The data generated by the digital twins will be directly shared in the DIG_IT platform and visualized by the final user.

Partner name	Data Sharing
ICCS	Data sharing will be decided in cooperation with end users, how and where the data will be displayed, and which partners will be granted access to. Hence, dissemination of the data will be decided in cooperation with the end users. No open data will be generated directly. The mines will decide what (or if) part of data will be made openly available.
LIBRA	All LIBRA generated and processed data are stored in the ICCS platform, so the same policies with the ICCS applies. Moreover, in the decision support system, data access will be regulated by the RBAC (Role Based Access Control). These roles will be created in agreement with the end-user needs.
STRATEGEM	Several one-to-one meetings took place between STRATAGEM and the mine owners to collect the data. The results will be shared with the consortium partners anytime in various formats; Word, Excel and LCA software files are the means which will be used to calculate, model and present the results of the study. The associated deliverable is a public document so it is open to the public however sensitive/confidential data are respected.
Titania	Titania has agreed to share our data with the partners in the Dig_IT project. However, our data cannot be shared outside the project or made open access without consent from Titania.
Marini Marmi	Marini has agreed to share the data collected on its use case with the partners in the Dig_IT project. Our data cannot be shared with third parties outside the project and/or made open access without consent from Marini.
TAPOJÄRVI	Tapojärvi has agreed to share our data with the partners in the Dig_IT project. Our data cannot be shared with third parties outside the project and/or made open access without consent from Tapojärvi.
ROTECH	All contents of the project website are public but permission for using the data is required according to Article 8.5.1 of Dig_IT Consortium Agreement.
BRUNEL	Not decided yet. To be decided in cooperation with end-user how and if/where the produced data will be used. Dissemination of the data will be decided upon in cooperation with the end-users. No open data will be generated directly.
SEI	The data remain inside the industrial PC and can be shared with specific security policy decided in accordance with the end user.
EUROCORE	No data will be shared with others.
IOTA	The data about the air and water quality will be provided over the public dashboard.

3.1.5 Preservation of the data during and after the end of the project

Table 7 Preservation of the data during and after the end of the project

Partner name	Data Sharing
SUBTERRA	We will have a copy of all the data in our company private service where it will be preserved after and at the end of the project.
SINTEF	Data for the evaluation and calibration of the sensor bundles will be stored on our own servers. When the sensors bundles are used by the end user, data will be managed by the Dig_IT-platform and not by SINTEF.
TAU	Mainly, the data will be stored on the edge server. In the early stages of the project, when the operation of the server is still uncertain, some backups of the data will be stored to local laptops and memory sticks. After the project, the server including the data is transferred to the Tapojärvi mine. TAU might keep some copies of the data.
CORE	The collected data through the Dig_IT-platform will be stored on our own servers for the duration required by Dig_IT project for the optimization and the calibration of the developed AI models. No preservation of end-user needs data after the project. Meta-data will be present only in PEDR as a deliverable.
ITAINNOVA	No data will be collected during nor after the project by ITAINNOVA.
ICCS	During the time span of the project data will be stored at the ICCS platform to facilitate the realization of the project's use cases, and their evaluation/impact in the mining industry. After the project's completion, the data handling (storage, deletion. etc.) will be decided by the respective mines and relevant project partners.
LIBRA	All LIBRA generated and processed data are stored in the ICCS platform, so the same policies with the ICCS applies.
STRATEGEM	Storage and backups of the relevant materials in local servers (PCs, laptops etc.) as the first level of storage and backup. All data and files will be self - archive and preserved in a Google Drive folder that has been created for the purposed of the project which will be up after the end of the Project and for 2 more years.
Titania	Titania is an end-user in this project and as such is a data owner and provider to the project. Thus, Titania will preserve the data related to Titania during and after the project.
Marini Marmi	Marini Marmi as end-user is a data owner and provider to the project. Data related to our case will be stored during and after the end of the project.
TAPOJÄRVI	Tapojärvi is an end-user in this project and as such is a data owner and provider to the project. Data related to our case will be stored by Tapojärvi during and after the project.
ROTECH	The data storage, backup, availability and long-term preservation are guaranteed by the web hosting service provider where a maximum 10 GB of storage have been preserved. The data preservation is guaranteed during the project and for a period of one year after the project.

Partner name	Data Sharing
BRUNEL	Data will be securely stored on Hard disks. Preservation of the data after the end of the project will be decided in cooperation with the other partners. [Partner name]: Please add your relevant contribution according to your work.
SEI	The data during the project is managed only by the researcher that is involved in the project. After the installation the data can be accessed only by the end user and from ICCS platform. The duration of retain data depends from the needs of end-user or from another partner.
EUROCORE	No data will be preserved during or after the project.
IOTA	The project will store data both locally and on the ledger. The local storage ensures a fast retrieval of the information, while the copy on the ledger ensures its immutability. Dig_IT will remove any personal information before sending the data to a blockchain. The data stored locally are encrypted through asymmetric encryption. For personal data, only the owner of the data will have the access to the data, upon authentication and authorization.

3.2 Dig_IT Datasets

Dig_IT datasets can be divided into two types of datasets. The first are the datasets gathered, processed and elaborated in the UCs while the rest are the other types of research data handled by partners.

3.2.1 Dig_IT Use Cases datasets

Table 8 summarises the datasets that will be generated over the project for each use case (Section 3.1), according to the categories indicated in Table 2. From Table 8, it is clear that not all kinds of data will be collected at every use case.

Table 8 Datasets for each use case

Research data	Dig_IT Use Cases				
	MARINI	LA PARRILLA	TITANIA	HANNUKAINEN	SOTKAMO
Geotechnical	Table 9	Table 21	Table 22	-	-
EHS & Sustainability	Table 10	-	Table 23	Table 34, Table 35, Table 36	Table 37

Research data	Dig_IT Use Cases				
	MARINI	LA PARRILLA	TITANIA	HANNUKAINEN	SOTKAMO
Situation & Biometric	Table 11, Table 12, Table 13, Table 15, Table 15	-	Table 24, Table 26, Table 27, Table 28	-	Table 38, Table 40, Table 41, Table 42, Table 43
Assets	Table 17	-	Table 29	-	Table 41
Predictive Maintenance Agents (PMA) predictions	Table 18		Table 30		Table 42
Predictive Operation System (POS) predictions	Table 18		Table 31		Table 43
Intelligent Toolbox for OHSE	Table 19		Table 32, Table 33		Table 44, Table 45

3.2.1.1 UC1 – Marini Marmi (underground quarry)

Marini Marmi srl (Italy) operate since 1993 in Ceppo di Gré underground quarry on the shore of Lake d’Iseo near Bergamo producing exclusive *Ceppo di Gré* and *Nuvolato di Gré* dimension stones used mainly for construction.

Through Dig_IT, MARINI aims to optimise safe operations and overall production control. A need for fluid dynamics monitoring, EHS and noise monitoring is provident, as well as maximising insights from assets regarding energy, waste, safety and scheduling. Table 4-10 describe the datasets of Marini Marmi UCs.

Table 9 Marini (UC 1) – Geotechnical Data – Geotechnical Dataset for Digital Twin

Dataset reference and name	Geotechnical Data for Digital Twin
Description	This dataset description represents MARINI in addressing their challenges in real time modelling of the underground exploitation. Detail of this description is subject to SUBTERRA Geotechnical Digital Twin. Dig_IT H2020: Human-centred Internet of Things Platform for the Sustainable Digital Mine of the Future project
Source	Geotechnical department of MARINI At the beginning of the project there is no monitoring system available for geotechnical purposes. The aim of Subterra is to test the UAV for underground monitoring.
Other partners contributing	SUBTERRA, Brunel
Personal data	No

Dataset reference and name	Geotechnical Data for Digital Twin
License	UAVDJI Inspire flights with RGB camera and multispectral sensor and image processing software
Embargo periods	N/A, we are still pending of a day to go to Marini and use the UAV flight
Legal framework(s)	N/A
Technology	UAV DJI Inspire flights with RGB camera and multispectral sensor
Standards and Metadata	N/A
Data sharing	UAV sensors:(Using USB File Transfer): Connectivity via local WIFI will be tested for real time acquisition. In an underground environment other options for connectivity will be analysed, by means of GSM or USB data download.
Data volume	Depending on data structure and data flow in the project
Data type and format	Excel, txt, word and pdf Files and jpg images
Update frequency	UAV: every 15 min flight the data can be downloaded. After battery charge another flight can be programmed.
Archiving and presentation	UAV sensors: high-definition set of RGB images. Point Clouds exportable to DXF files. Processing software to create Digital Model of the Ground.
Responsibilities	SUBTERRA is in charge of the Geotechnical Digital Twin in the selected sites. In this regard, data acquisition from the described sensors during WP2 will be carried out. Further updates of the geometrical and geotechnical model will be accomplished by means of introducing that data in the calculation code.
Resources	UAV Sensors-to prevent swelling, the battery automatically discharges to below 65% of its total power capacity if it is not used for more than 10 days. It takes approximately 2 days to discharge a fully charged battery's power level to 65%. It is normal to feel moderate heat emanating from the battery during this process. Discharge thresholds can be modified using the DJI GO app.
Additional services and provisions	N/A
Ethical aspects	N/A
Other	N/A

Table 10 Marini (UC 1) – EHS & Sustainability data - OHSE Dataset

Dataset reference and name	OHSE Data
Description	OHSE dataset – generic for mines. Sensors for surveying OHSE. Potential audience of interest: End users.
Source	Sensors
Other partners contributing	End user: Marini, SUBTERRA, Brunel
Personal data	No
Licence	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	Not identified yet.
Standards and Metadata	From D1.5: Sensors for online measurement of OSH-parameters: Noise

Dataset reference and name	OHSE Data
	<p>Diesel exhaust (NO₂ and CO)</p> <p>OSH- parameters: Flow charts of data flow is described in fig. 3-1, 3-3 in D1.5</p> <p>Data will be created from scratch (on-line measurements and at-line measurements) and statistical treated data based on these.</p> <p>Metadata protocol will be decided by other partners in Dig IT.</p>
Data sharing	Not decided yet. To be decided in cooperation with the end user how and where the data will be displayed at the end user. Dissemination of the data will be decided upon in cooperation with the end users. No open data will be generated directly.
Data volume	Depending on data structure and data flow in the project.
Data type and format	N/A
Update frequency	Suggested in the flowcharts.
Archiving and presentation	Treated by other partners.
Responsibilities	N/A
Resources	N/A
Additional services and provisions	Treated by other partners
Ethical aspects	No personal data. No ethical aspects identified.
Other	

Table 11 Marini (UC 1) – Situational awareness & biometrical data - Stress & Fatigue Monitoring Dataset

Dataset reference and name	Stress & Fatigue Monitoring Data
Description	Data from users' wristbands, which is used to determine the overall stress and fatigue levels users. Potential audience of interest: End users.
Source	Heart rate sensor – PPG Accelerometers/gyroscopes Skin Temperature GalvanicSkinResponse Sensor - GSR (“micro-Mho (μM)” or “micro-Siemens (μS)”)
Other partners contributing	End users: Marini, SUBTERRA, ICCS
Personal data	Yes
License	Not to be disclosed outside of consortium without consent of user
Embargo periods	N/A
Legal framework(s)	GDPR, article 9 (2) GDPR
Technology	Wristband developed by ICCS, please refer to Deliverable 2.4.
Standards and Metadata	From D1.8: Examine collapse, muscle cramps, heart palpitation, carelessness, accidents, mental health problems, chronic health issues
Data sharing	Not decided yet. To be decided in cooperation with end user how and where the data will be displayed at the end user. Dissemination of the data will be decided in cooperation with the end users. No open data will be generated directly.
Data volume	Depending on data structure and data flow in the project.
Data type and format	Proposed: JSON

Dataset reference and name	Stress & Fatigue Monitoring Data
Update frequency	Not decided yet. It will be decided based on the smart garment capabilities/limitations (e.g., device battery drainage, processing requirements) and use case needs.
Archiving and presentation	Data will flow from the end devices to the ICCS platform that will be delivered in WP3 and detailed in D3.1 (Dig_IT IIoT platform architecture design). The accumulated data will be stored at the platform and exploited by other partners (e.g., LIBRA, CORE) for analytics services and visualization. The data preservation policy will be decided in collaboration with the end users (Mines) and use case requirements.
Responsibilities	All partners present on the data flow including data production, transmission, processing and consumption.
Resources	<ul style="list-style-type: none"> • Sensors that collect/measure the biometrics data. • Bluetooth communication module that interconnects the wristband to the smart garment. • Smart garment main processing unit that collects sensed data (format, prepare and send over WiFi). • WiFi connectivity to expose the data outside the mines area(s) towards the Dig_IT partners and platforms. <p>Please refer to D2.4 for more details regarding the smart garment components (wristband, headset, garment).</p> <p>Data maintenance will be decided by the end users and partners involved in data acquisition, processing and storage as will be detailed in WP3 and WP5.</p>
Additional services and provisions	N/A
Ethical aspects	Biometric data. Processing of biometric data for the purpose of uniquely identifying natural persons is prohibited. The user's heart rate information and identification information are sensitive and must be compliant with GDPR rules. Data integrity and security needs to be ensured.
Other	N/A

Table 12 Marini (UC 1) – Situational awareness & biometrical data - Location Dataset

Dataset reference and name	Location Data
Description	Location of users in the mine
Source	Wearable device
Other partners contributing	End users: Marini, SUBTERRA, ICCS
Personal data	Yes
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	GDPR
Technology	The technology that will be exploited for the localization service will be based on ultra-wideband (UWB) technology. Please refer to D2.4 for more details.
Standards and Metadata	Not yet available

Dataset reference and name	Location Data
Data sharing	Not decided yet. To be decided in cooperation with end user how and where the data will be displayed at the end user. Dissemination of the data will be decided in cooperation with the end users. No open data will be generated directly.
Data volume	Depending on data structure and data flow in the project.
Data type and format	Proposed: JSON file indicating x,y,z coordinates of the person within the mine area.
Update frequency	Not decided yet. It will be decided based on the smart garment capabilities/limitations (e.g., device battery drainage, processing requirements) and use case needs.
Archiving and presentation	Real time stream that does not require storage. ICCS generates the location info that is exploited by the respective task in WP5.
Responsibilities	ICCS and task partners in WP5.
Resources	<ul style="list-style-type: none"> • A network of UWB sensors placed within the mine. • Smart garment to transmit the calculated position based on the UWB metrics to the ICCS platform and task partners of WP5 <p>Please refer to D2.4 for more details.</p>
Additional services and provisions	N/A
Ethical aspects	Personal data. The user's location information is sensitive and must be compliant with GDPR rules. Data integrity and security needs to be ensured
Other	N/A

Table 13 Marini (UC 1) – Situational awareness & biometrical data - Wearable OHSE Dataset

Dataset reference and name	OHSE Wearable Data
Description	OHSE dataset – generic for mines. Sensors for surveying OHSE. Potential audience of interest: End users.
Source	Ambient temperature sensor Humidity Sensor Air quality sensors (e.g., CO, CH ₄ , NH ₃ , NO _x , CO ₂) Noise level sensors
Other partners contributing	End user: Marini, SUBTERRA, ICCS
Personal data	No
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	Detailed information for the sensor types used and means of connectivity among them are detailed in D2.4.
Standards and Metadata	Please refer to D1.5 and D2.4.
Data sharing	Not decided yet. To be decided in cooperation with end user how and where the data will be displayed at the end user. Dissemination of the data will be decided upon in cooperation with the end users. No open data will be generated directly.
Data volume	Depending on data structure and data flow in the project.

Dataset reference and name	OHSE Wearable Data
Data type and format	Proposed: JSON
Update frequency	Not decided yet. It will be decided based on the smart garment capabilities/limitations (e.g., device battery drainage, processing requirements) and use case needs.
Archiving and presentation	Data will flow from the end devices to the ICCS platform that will be delivered in WP3 and detailed in D3.1 (Dig_IT IIoT platform architecture design). The accumulated data will be stored at the platform and exploited by other partners (e.g., LIBRA, CORE) for analytics services and visualization. The data preservation policy will be decided in collaboration with the end users (Mines) and use case requirements.
Responsibilities	ICCS and task partners in WP5.
Resources	<ul style="list-style-type: none"> • OHSE sensors. • Smart garment main processing unit that collects sensed data (format, prepare and send over WiFi). • WiFi connectivity to expose the data outside the mines area(s) towards the Dig_IT partners and platforms. <p>Please refer to D2.4 for more details regarding the smart garment components (wristband, headset, garment).</p> <p>Data maintenance will be decided by the end users and partners involved in data acquisition, processing and storage as will be detailed in WP3 and WP5.</p>
Additional services and provisions	N/A
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 14 Marini (UC 1) – Digital Twin Air Quality Data

Dataset reference and name	Digital twin air quality data (Marini Marmi)
Description	<ul style="list-style-type: none"> • Air quality properties (Residence time, Air change efficiency, Temperature...) inside the quarry. • Used by the final user for information and decision supporting.
Source	Digital twin (virtual sensor) embedded into the DIG_IT platform.
Other partners contributing	Sinteff: Data acquisition. Schneider: Data aggregator. LIBRA: Decision support system. Marini: Final user
Personal data	NO
License	N/A
Embargo periods	N/A
Legal framework(s)	N/A
Technology	Digital twin generated by using ITAINNOVA in house CFD model (OpenFOAM based) and reduced order model library (Twinkle)

Dataset reference and name	Digital twin air quality data (Marini Marmi)
Standards and Metadata	Real time air quality prediction inside the mine is generated by the digital twin. Properties as temperature, air change efficiency or the time that air has remained inside the mine are shown for different mine zones as well as a distribution contour plot. The data can be used to generate risk maps and alerts that support the final user for optimization and improvement of air quality during working hours.
Data sharing	Data is generated inside the DIG_IT platform and shown in the platform DSS, which is accessible by the final user.
Data volume	< 1Mb.
Data type and format	Image formatting and text or csv files.
Update frequency	The frequency will depend on the frequency of the data collected by the physical sensor (weather station), which is the input for the digital twin (virtual sensor)
Archiving and presentation	Data is presented in the platform through the user interface (DSS) developed by LIBRA.
Responsibilities	N/A
Resources	N/A
Additional services and provisions	Other partners in charge.
Ethical aspects	No personal data. No ethical aspects identified.
Other	

Table 15 Marini (UC 1) – Situational awareness & biometrical data - Smart Garment Command & Control Dataset

Dataset reference and name	Smart Garment Command & Control Data
Description	Data for controlling devices & commands from users.
Source	Smart earplug/headset
Other partners contributing	End user: Marini, SUBTERRA, ICCS
Personal data	No
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	Please refer to D2.4.
Standards and Metadata	Not decided yet.
Data sharing	Not decided yet. To be decided in cooperation with end user how and where the data will be displayed at the end user. Dissemination of the data will be decided upon in cooperation with the end users. No open data will be generated directly.
Data volume	Keyword that is detected.
Data type and format	Proposed: JSON
Update frequency	Event driven (triggered by the smart garment user)
Archiving and presentation	Detected keyword will reach the ICCS platform and potential trigger alert back to the user. The data preservation policy will be decided in collaboration with the end users (Mines) and use case requirements.
Responsibilities	ICCS and end users.
Resources	Headset subsystem as described in D2.4
Additional services and provisions	N/A
Ethical aspects	No personal data. No ethical aspects identified.

Dataset reference and name	Smart Garment Command & Control Data
Other	N/A

Table 16 Marini (UC 1) – Assets data

Dataset reference and name	Marini ² Diamond Gangsaw Data
Description	Measurement data from the diamond gangsaw - PTM Kuasar 900. Process/Machine data from the cutting process of blocks into slabs for developing POS tool.
Source	PTM Kuasar 900 control system, Core System
Other partners contributing	CORE, Marini, Brunel
Personal data	No
Licence	Proprietary License
Embargo periods	N/A
Legal framework(s)	Dig IT NDA
Technology	N/A
Standards and Metadata	N/A
Data sharing	No open data will be generated directly.
Data volume	Depending on data structure and data update frequency.
Data type and format	JSON schema
Update frequency	N/A
Archiving and presentation	N/A
Responsibilities	CORE, Marini
Resources	Marini, AWS/On-Premises Hybrid cloud.
Additional services and provisions	Time-series DB, Time-series virtualization dashboard, Containerized/Orchestration technologies.
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 17 Marini (UC 1) – PMA Data

Dataset reference and name	PMA Data
Description	Future States and Failures in RT. Individual assets health (machine, tools and vehicles) with related historical data (ID: PMA_001)
Source	CORE, MARINI
Other partners contributing	MARINI
Personal data	No
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	REST API

² MARINI, BASED ON DELIVERABLE D1.7 (P18) AT TABLE 5, USER REQUIREMENT ID P001, UDA IS NOT A MAJOR CONCERN FOR THEM. HOWEVER, THE PMA WILL BE APPLIED FOR THE FUTURE STATE AND FAILURE OF THE ASSET.

Dataset reference and name	PMA Data
Standards and Metadata	Neural network (Encoder – Decoder architecture) consuming multi-variate time-series from the sensor data that are placed on each asset.
Data sharing	LIBRA No open data will be generated directly.
Data volume	N/A
Data type and format	JSON schema
Update frequency	N/A
Archiving and presentation	Dig_IT IIoT platform, Dashboard (LIBRA).
Responsibilities	Dig_IT IIoT platform
Resources	Asset Data (MARINI), Maintenance data (CORE)
Additional services/ provisions	AWS/On-Premises Hybrid cloud, Time-series DB, Time-series virtualization dashboard
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 18 Marini (UC 1) – POS Data

Dataset reference and name	Predictive Operations System (POS) predictions
Description	Energy Consumption (ID: POS_003)
Source	MARINI
Other partners contributing	CORE, MARINI
Personal data	No
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	REST API
Standards and Metadata	Neural network (LSTM and/or Transformer architectures) consuming multi-variate time-series sensor data placed on the asset
Data sharing	LIBRA No open data will be generated directly.
Data volume	N/A
Data type and format	JSON schema
Update frequency	N/A
Archiving and presentation	Dig_IT IIoT platform, Dashboard (LIBRA).
Responsibilities	Dig_IT IIoT platform
Resources	Asset Data (MARINI)
Additional services/ provisions	AWS/On-Premises Hybrid cloud, Time-series DB, Time-series virtualization dashboard
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 19 Marini (UC 1) – Intelligent Toolbox for OHSE Data

Dataset reference and name	Intelligent Toolbox for OHSE
Description	Bio Signals Analytics & Anomaly Detection
Source	Dig_IT IIoT platform
Other contributing partners	ICCS,CORE, MARINI, SEI
Personal data	Yes
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	GDPR, Article 9 GDPR
Technology	N/A
Standards and Metadata	Neural network (Encoder – Decoder architecture) consuming multi-variate time-series from the sensor data that are smart garment.
Data sharing	LIBRA No open data will be generated directly
Data volume	N/A
Data type and format	JSON schema
Update frequency	N/A
Archiving and presentation	Dig_IT IIoT platform, Dashboard (LIBRA).
Responsibilities	Dig_IT IIoT platform
Resources	Dig_IT IIoT platform
Additional services/provisions	AWS/On-Premises Hybrid cloud, Time-series DB, Time-series virtualization dashboard
Ethical aspects	Biometric data. Processing of biometric data for the purpose of uniquely identifying natural persons is prohibited. The user's heart rate information and identification information are sensitive and must be compliant with GDPR rules. Data integrity and security needs to be ensured.
Other	N/A

3.2.1.2 UC2 – La Parrilla mine (open pit)

La Parrilla mine (Spain) is located in Caceres province (Extremadura). It is owned by *W Resources*, a European mining company focusing on long term tungsten production owning five mines in Spain and Portugal.

La Parrilla aims for an advanced monitoring system to follow geometrical and geotechnical conditions that change along the excavation progress, emphasising instability-prone areas is critical for establishing optimal production planning. Monitoring would include significant ground parameters i.e.: horizontal and vertical displacements of benches, spoil heaps and retaining walls; pore pressure evolution in pits and spoils; total pressure in significant locations. La Parrilla intends to prevent instability episodes and reduce uncertainties in the infrastructure behaviour. They could highly benefit from a predictive feature of the system allowing to exploit the ore under safer conditions. Moreover, intelligent insights for making decisions regarding planning of production, geometry of the

pit, and early alerts would optimise their decision-making time both on site and at the office. Table 11 describes the dataset La Parrilla UC.

Table 20 La Parrilla (UC 2) – Geotechnical Data – Geotechnical Dataset for Digital Twin

Dataset reference and name	Geotechnical Data for Digital Twin
Description	This dataset description represents LA PARRILLA (Western Resources mining company) in addressing their challenges in real time modelling of their tailing dam barrier. Detail of this description is subject to SUBTERRA Geotechnical Digital Twin. Dig_IT H2020: Human-centred Internet of Things Platform for the Sustainable Digital Mine of the Future project
Source	Geotechnical department of LA PARRILLA Vibrating Wire Piezometer Dynamic Inclinator Data from topographical station UAV DJI Inspire flights with RGB camera and multispectral sensor Piezometer Sensor: 4500AR Piezometer via LC2 Series 4-Channel Data loggers Variables: Data Logger ID, Date (Y/M/D), Time (Hr/Min/Sec), Battery Voltage (V), Internal Temperature (°C), Vibrating Sensor Reading (deg), Sensor Temperature (°C), Array Number
Other partners contributing	SUBTERRA, LA PARRILLA
Personal data	No
License	Radar (Product License): nIDS GeoRadar, Hexagon USA. – FPM360 License Package which covers alarming option, Data Analysis Tools, Automatic Atmospheric Correction, Multiscale Processing, 3D Interface, Geodetic Data (GeoMoS, T4D), API, Server/Client, Planning Tool, FPM360 and True Vector. The TILT-5x dynamic Tilt / Inclinator family is a series of high performance, high resolution three axis digital dynamic or gyro compensated inclinometers that use the latest miniature technology in the form of accelerometer and gyroscope MEMS sensors. It is contained within a robust aluminium housing and several interface options such as USB, RS232, RS485, RS422, and 3.3 VTTL UART. Piezometer (Product License): Geokon Inc, USA covers the Piezometer sensors, Data Loggers and LogView Software.
Embargo periods	N/A
Legal framework(s)	N/A
Technology	Details of the radar system employed and other sensors for data acquisition (piezometers, modeling software) Piezometer: Device - Piezometer Sensor, Data Logger. Communication Protocol - RS232, and USB 2.0 connection. Software - LogView Software. UAV DJI Inspire flights with RGB camera and multispectral sensor
Standards and Metadata	N/A
Data sharing	Piezometer: (Using USB File Transfer): engineer will collect data manually from data logger and transfer to LogView software for data analysis. Remark: LogView software support real-time data for further expansion. Inclinometer: the TILT-5x Dynamic Tilt series uses two LEDs that show the communication status of sensors while receiving data, sending data or computing offline.

Dataset reference and name	Geotechnical Data for Digital Twin
	UAV sensors:(Using USB File Transfer): Connectivity via local WIFI will be tested for real time acquisition
Data volume	N/A
Data type and format	Excel, txt, word and pdf Files and jpg images
Update frequency	Piezometer: every 4 hours, 6 times a day (offline) Inclinometer: every 24 hours, one measure unless other set up is required.
Archiving and presentation	Piezometer - inclinometer: text and Graphic Views for static and dynamic display of data using LogView software. LogView Software simplifies the task of configuration, communication, monitoring, data collection and data reduction using the data loggers. UAV sensors: high-definition set of RGB images. Point Clouds exportable to DXF files. Processing software to create Digital Model of the Ground.
Responsibilities	SUBTERRA is in charge of the Geotechnical Digital Twin in the selected sites. In this regard, data acquisition from the described sensors during WP2 will be carried out. Further updates of the geometrical and geotechnical model will be accomplished by means of introducing that data in the calculation code.
Resources	Piezometer - Piezometers records readings of digits (used to calculate water pressure), battery voltage, internal temperature and sensor temperature. Each borehole has 2-3 piezometers recording data at different levels, all transmitted to a common data-logger (1 data-logger per borehole). The data is recorded by a logger, and the user has to manually visit each piezometer to download and view the data (i.e., used for back-analysis). Inclinometer - While inclinometers measure the angle of an object with respect to the force of gravity, any external acceleration (any type of motion such as movement, vibration, etc.) will introduce errors in tilt measurements. To overcome this problem, it is possible to add a gyroscope in addition to the accelerometer. Such a system combined with algorithms to process both signals (Data from accelerometer and gyroscope) in order to get the best value out of each sensor, is called a Dynamic Inclinometer. UAV Sensors-to prevent swelling, the battery automatically discharges to below 65% of its total power capacity if it is not used for more than 10 days. It takes approximately 2 days to discharge a fully charged battery's power level to 65%. It is normal to feel moderate heat emanating from the battery during this process. Discharge thresholds can be modified using the DJI GOapp.
Additional services and provisions	N/A
Ethical aspects	N/A
Other	N/A

3.2.1.3 UC3 – Titania AS (open pit)

Titania AS (Norway) is a part of Kronos Worldwide Inc. (US) producer and marketer of titanium dioxide (TiO₂) pigments. They operate on a 2,7km by 0,6km, 320m deep open-pit mine located in

Hauge I Dalane, southern Norway, where ilmenite-norite is extracted to produce ilmenite concentrate (anorthosite as waste rock).

With Dig_IT, Titania aims to improve their monitoring and control of water discharge control from the mine, processing plant and land fill. Moreover, they aim in strengthening the industry social presence within the local community through establishing a relationship of trust regarding environmental compliance.

Titania aims to take slope monitoring one step further in integrating data from various sources and sensors into one unified system that can be updated in real time. Titania aspires for timely insights regarding operation and maintenance needs of their critical assets i.e., the optimal time for performing maintenance activities in order to prevent unexpected delays in production.

Therefore, Titania aspires for an overall integrated agile system to obtain critical information on time for scheduling production regarding asset maintenance, energy consumption, quality of production, environmental performance and overall operation, that is relevant for the whole value chain of the mine.

Table 21 Titania (UC 3) – Geotechnical Data – Geotechnical Dataset for Digital Twin

Dataset reference and name	Geotechnical Data for Digital Twin
Description	This dataset description represents TITANIA AS mining company in addressing their challenges in real time modelling of one of the pit slopes. Detail of this description is subject to SUBTERRA Geotechnical Digital Twin. Dig_IT H2020: Human-centred Internet of Things Platform for the Sustainable Digital Mine of the Future project
Source	Geotechnical department of TITANIA Radar Sensor: Rover_2015, Rover_2018 via Wi-Fi Ethernet Switch Variables: Acceleration, Amplitude, Displacement (mm/hr), Velocity, Digital Terrain Model, Humidity (%), Pressure (mbar), Temperature (°C), Battery Voltage (V), Solar Panel (W). Piezometer Sensor: 4500AR Piezometer via LC2 Series 4-Channel Data loggers Variables: Data Logger ID, Date (Y/M/D), Time (Hr/Min/Sec), Battery Voltage (V), Internal Temperature (°C), Vibrating Sensor Reading (deg), Sensor Temperature (°C), Array Number
Other partners contributing	TITANIA, Brunel
Personal data	No
License	Radar (Product Licence): IDS GeoRadar, Hexagon USA. – FPM360 Licence Package which covers alarming option, Data Analysis Tools, Automatic Atmospheric Correction, Multiscale Processing, 3D Interface, Geodetic Data (GeoMoS, T4D), API, Server/Client, Planning Tool, FPM360 and True Vector. Piezometer (Product Licence): Geokon Inc, USA covers the Piezometer sensors, Data Loggers and LogView Software.
Embargo periods	N/A

Dataset reference and name	Geotechnical Data for Digital Twin
Legal framework(s)	N/A
Technology	<p>Details of the radar system employed and other sensors for data acquisition (piezometers, modelling software)</p> <p>Radar:</p> <p>Device - Sensor Pointing Camera, Eagle-Vision Camera, Power Supply Controller Board, Weather Station. Server - Guardian Master FPM and Data Server, Guardian Processor FPM 1 and 2, Watchdog, Guardian Client FPM workstations. Communication Protocol - Fiber Optic, Wi-Fi Radio (AP), Wi-Fi Antenna (Stn), Ethernet Hub. Software - IBIS Guardian Software.</p> <p>Piezometer:</p> <p>Device - Piezometer Sensor, Data Logger. Communication Protocol - RS232, and USB 2.0 connection. Software - LogView Software.</p>
Standards and Metadata	N/A
Data sharing	<p>Radar: (Using FPM360 architecture). Data is transmitted From IBS Units to IBIS Guardian FPM Processor for data processing. Then it transferred to IBIS Guardian FPM Master & Data server for Global data visualisation, analysis alarming and database storage. Other details -</p> <ol style="list-style-type: none"> 1) Alarming capabilities - Alarm generation with user-defined levels and multiple alarm criteria based upon: Displacement, Velocity, Acceleration, Inverse Velocity parameters. 2) Data Analysis Tools - Extensive tools for data analysis and interpretation, including extraction of time series from user-defined areas, inverse of velocity charts, import of digital layers, movie tool. <p>Piezometer: (Using USB File Transfer): engineer will collect data manually from data logger and transfer to LogView software for data analysis. Remark: LogView software supports real-time data for further expansion.</p>
Data volume	N/A
Data type and format	Excel, word and pdf Files and jpg images
Update frequency	Radar: every 4 minutes, 360 times a day (online) Piezometer: every 4 hours, 6 times a day (offline)
Archiving and presentation	<p>Radar: 3D interactive data visualisation and fully geo-referenced radar data for export to mine planning software and GIS. Using IBIS Guardian software, an extremely powerful and feature-rich monitoring platform for managing slope stability risks in open-pit mining operations.</p> <p>Piezometer: text and Graphic Views for static and dynamic display of data using LogView software. LogView Software simplifies the task of configuration, communication, monitoring, data collection and data reduction using the data loggers.</p>
Responsibilities	SUBTERRA is in charge of the Geotechnical Digital Twin in the selected sites. In this regard, data acquisition from the described sensors during WP2 will be carried out. Further updates of the geometrical and geotechnical model will be accomplished by means of introducing that data in the calculation code.
Resources	Radar - The radars record a variety of data. They detect displacement (pit-wall movement) from measuring the phase-

Dataset reference and name	Geotechnical Data for Digital Twin
	<p>shift between the signal sent from the transmitter and received by the receiver. A classification of the signal strength and sensitivity is also made for each recorded pixel. One of the radars is also equipped with a weather station measuring humidity, precipitation, pressure and temperature. All these data are sent from the radars to the master computer in the office, for real-time data. The radars also record system status internally (system batteries, generator status, camera, generator fuel etc.).</p> <p>Piezometer - Piezometer's records readings of digits (used to calculate water pressure), battery voltage, internal temperature and sensor temperature. Each borehole has 2-3 piezometers recording data at different levels, all transmitted to a common data-logger (1 data-logger per borehole). The data is recorded by a logger, and the user has to manually visit each piezometer to download and view the data (i.e., used for back-analysis).</p>
Additional services and provisions	N/A
Ethical aspects	N/A
Other	N/A

Table 22 Titania (UC 3) – EHS & Sustainability data - OHSE Dataset

Dataset reference and name	OHSE Data
Description	OHSE dataset – generic for mines. Sensors for surveying OHSE. Potential audience of interest: End users.
Source	Sensors
Other partners contributing	End user: Titania, Brunel
Personal data	No
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	Not identified yet.
Standards and Metadata	<p>From D1.5:</p> <p>Online measurement of water parameters:</p> <p>Flow measurements</p> <p>Suspended material (turbidity)</p> <p>Ni, Cu</p> <p>Sensors for online measurement of OSH-parameters:</p> <p>Noise</p> <p>Dust</p> <p>Diesel exhaust (NO₂ and CO)</p> <p>OSH - parameters: Flow charts of data flow is described in fig. 3-1, 3-2, 3-3 in D1.5.</p> <p>Water parameters: Flow charts of data flow is described in fig. 4-2 and 4-3 in D1.5</p> <p>Data will be created from scratch (on-line measurements and at-line measurements) and statistical treated data based on these.</p> <p>Metadata protocol will be decided by other partners in Dig_IT.</p>

Dataset reference and name	OHSE Data
Data sharing	Not decided yet. To be decided in cooperation with end user how and where the data will be displayed at the end user. Dissemination of the data will be decided upon in cooperation with the end users. No open data will be generated directly.
Data volume	Depending on data structure and data flow in the project.
Data type and format	N/A
Update frequency	Suggested in the flowcharts.
Archiving and presentation	Treated by other partners.
Responsibilities	N/A
Resources	N/A
Additional services and provisions	Treated by other partners
Ethical aspects	No personal data. No ethical aspects identified.
Other	

Table 23 Titania (UC 3) – Situational awareness & biometrical data - Location Dataset

Dataset reference and name	Location Data
Description	Location of users in the mine
Source	Wearable device
Other partners contributing	End users: Titania, ICCS
Personal data	Yes
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	GDPR
Technology	The technology that will be exploited for the localization service will be based on ultra-wideband (UWB) technology. Please refer to D2.4 for more details.
Standards and Metadata	Not yet available
Data sharing	Not decided yet. To be decided in cooperation with end user how and where the data will be displayed at the end user. Dissemination of the data will be decided in cooperation with the end users. No open data will be generated directly.
Data volume	Depending on data structure and data flow in the project.
Data type and format	Proposed: JSON file indicating x,y,z coordinates of the person within the mine area.
Update frequency	Not decided yet. It will be decided based on the smart garment capabilities/limitations (e.g., device battery drainage, processing requirements) and use case needs.
Archiving and presentation	Real time stream that does not require storage. ICCS generates the location info that is exploited by the respective task in WP5.
Responsibilities	ICCS and task partners in WP5.
Resources	<ul style="list-style-type: none"> • A network of UWB sensors placed within the mine. • Smart garment to transmit the calculated position based on the UWB metrics to the ICCS platform and task partners of WP5 <p>Please refer to D2.4 for more details.</p>
Additional services and provisions	N/A

Dataset reference and name	Location Data
Ethical aspects	Personal data. The user's location information is sensitive and must be compliant with GDPR rules. Data integrity and security needs to be ensured
Other	N/A

Table 24 Titania (UC 3) – Situational awareness & biometrical data - Wearable OHSE Dataset

Dataset reference and name	OHSE Wearable Data
Description	OHSE dataset – generic for mines. Sensors for surveying OHSE. Potential audience of interest: End users.
Source	Ambient temperature sensor Humidity Sensor Air quality sensors (e.g., CO, CH ₄ , NH ₃ , NO _x , CO ₂) Noise level sensors
Other partners contributing	End user: Titania, ICCS
Personal data	No
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	Detailed information for the sensor types used and means of connectivity among them are detailed in D2.4.
Standards and Metadata	Please refer to D1.5 and D2.4.
Data sharing	Not decided yet. To be decided in cooperation with end user how and where the data will be displayed at the end user. Dissemination of the data will be decided upon in cooperation with the end users. No open data will be generated directly.
Data volume	Depending on data structure and data flow in the project.
Data type and format	Proposed: JSON
Update frequency	Not decided yet. It will be decided based on the smart garment capabilities/limitations (e.g., device battery drainage, processing requirements) and use case needs.
Archiving and presentation	Data will flow from the end devices to the ICCS platform that will be delivered in WP3 and detailed in D3.1 (Dig_IT IIoT platform architecture design). The accumulated data will be stored at the platform and exploited by other partners (e.g., LIBRA, CORE) for analytics services and visualization. The data preservation policy will be decided in collaboration with the end users (Mines) and use case requirements.
Responsibilities	ICCS and task partners in WP5.
Resources	<ul style="list-style-type: none"> • OHSE sensors. • Smart garment main processing unit that collects sensed data (format, prepare and send over WiFi). • WiFi connectivity to expose the data outside the mines area(s) towards the Dig_IT partners and platforms. <p>Please refer to D2.4 for more details regarding the smart garment components (wristband, headset, garment).</p>

Dataset reference and name	OHSE Wearable Data
	Data maintenance will be decided by the end users and partners involved in data acquisition, processing and storage as will be detailed in WP3 and WP5.
Additional services and provisions	N/A
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 25 Titania (UC 3) – Digital Twin Air Quality Data

Dataset reference and name	Digital twin air quality data (Titania)
Description	Water quality properties (Pollutant's concentration) assessment, used by the final user for information, decision supporting and optimization tasks.
Source	Digital twins (virtual sensors) embedded into the DIG_IT platform.
Other partners contributing	Sinteff: Data acquisition. Schneider: Data aggregator. LIBRA: Decision support system. Sotkamo: Final user
Personal data	NO
Dataset reference and name	Digital twin air quality data (Titania)
License	N/A
Embargo periods	N/A
Legal framework(s)	N/A
Technology	Digital twin generated by using ITAINNOVA in house dynamic models and reduced order model library (Twinkle)
Standards and Metadata	Real time assessment of the water quality is generated by the digital twin. The data can be used to generate risk maps and alerts that support the final user for optimization and improvement of water quality.
Data sharing	Data is generated inside the DIG_IT platform and shown in the platform DSS, which is accessible by the final user.
Data volume	< 1Mb.
Data type and format	Image formatting and text or csv files.
Update frequency	To be defined
Archiving and presentation	Data is presented in the platform through the web user interface developed by LIBRA.
Responsibilities	N/A
Resources	N/A
Additional services and provisions	Other partners in charge.
Ethical aspects	No personal data. No ethical aspects identified.
Other	

Table 26 Titania (UC 3) – Situational awareness & biometrical data - Smart Garment Command & Control Dataset

Dataset reference and name	Smart Garment Command & Control Data
Description	Data for controlling devices & commands from users.
Source	Smart earplug/headset
Other partners contributing	End user: Titania, SUBTERRA, TAU
Personal data	No

Dataset reference and name	Smart Garment Command & Control Data
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	Please refer to D2.4.
Standards and Metadata	Not decided yet.
Data sharing	Not decided yet. To be decided in cooperation with end user how and where the data will be displayed at the end user. Dissemination of the data will be decided upon in cooperation with the end users. No open data will be generated directly.
Data volume	Keyword that is detected.
Data type and format	Proposed: JSON
Update frequency	Event driven (triggered by the smart garment user)
Archiving and presentation	Detected keyword will reach the ICCS platform and potential trigger alert back to the user. The data preservation policy will be decided in collaboration with the end users (Mines) and use case requirements.
Responsibilities	ICCS and end users.
Resources	Headset subsystem as described in D2.4
Additional services and provisions	N/A
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 27 Titania (UC 3) – Assets data - Trucks Dataset

Dataset reference and name	TITANIA Trucks Data
Description	Measurement data from Titania's trucks for developing machine learning models and digital twins and as well as performing model verification.
Source	Sensors and event logs.
Other partners contributing	TAU, Titania, Brunel
Personal data	N/A
Licence	N/A
Embargo periods	N/A
Legal framework(s)	Dig_IT NDA.
Technology	.csv-files currently, research on APIs ongoing.
Standards and Metadata	Data is created by the on-board sensors of the trucks working in Titania's mine. Currently only historical data is accessed, real time transfer will be considered later. Once the initial dataset is complete, metadata description of it will be added.
Data sharing	The data can be shared with the Dig_IT research partners. Currently the plans for sharing it are not finalised, as the dataset is still being collected by Titania.
Data volume	Unknown, data collection is still in progress with some technical difficulties on Titania's end. Ideally some gigabytes, spanning at least a year of measurements.
Data type and format	Time series data of various frequencies as .csv-files. Categorised as events, data logger data, payload data and GPS data.
Update frequency	Unknown.
Archiving and presentation	The data will be initially archived and backed up on secure TAU hard drives, and later shared with other research

Dataset reference and name	TITANIA Trucks Data
	partners through means chosen together. Data will not be archived beyond the duration of the project.
Responsibilities	The data will be managed by the TAU research group.
Resources	Titania IT department, Titania project partners, TAU research group.
Additional services and provisions	Uncertain at this time, as the collection of the dataset is still on going.
Ethical aspects	N/A
Other	N/A

Table 28 Titania (UC 3) – PMA Data

Dataset reference and name	PMA Data
Description	Future States and Failures in RT. Individual assets health (machine, tools and vehicles) with related historical data (ID: PMA_001) Predictive maintenance with regards to emissions (ID: PMA_002)
Source	Dig_IT IIoT platform, TAU
Other partners contributing	CORE,TAU, TITANIA, SEI
Personal data	No
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	N/A
Standards and Metadata	Neural network (Encoder – Decoder architecture) consuming multi-variate time-series from the sensor data that are placed on each asset. Neural network (LSTM and/or Transformer architectures) consuming multi-variate time-series sensor data placed on the asset
Data sharing	LIBRA No open data will be generated directly.
Data volume	N/A
Data type and format	JSON schema
Update frequency	N/A
Archiving and presentation	Dig_IT IIoT platform, Dashboard (LIBRA).
Responsibilities	Dig_IT IIoT platform
Resources	RT-DT (TAU), Dig_IT IIoT platform, Maintenance data (TAU)
Additional services/ provisions	AWS/On-Premises Hybrid cloud, Time-series DB, Time-series virtualization dashboard
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 29 Titania (UC 3) – POS Data

Dataset reference and name	POS Data
Description	Downtime prediction per asset (ID: POS_001) Energy Consumption (ID: POS_003)
Source	Dig_IT IIoT platform, TAU
Other partners contributing	CORE,TAU, TITANIA, SEI
Personal data	No
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	N/A
Standards and Metadata	Neural network (LSTM and/or Transformer architectures) consuming multi-variate time-series sensor data placed on the asset
Data sharing	LIBRA No open data will be generated directly.
Data volume	N/A
Data type and format	JSON schema
Update frequency	N/A
Archiving and presentation	Dig_IT IIoT platform, Dashboard (LIBRA).
Responsibilities	Dig_IT IIoT platform
Resources	RT-DT (TAU), Dig_IT IIoT platform, Maintenance data (TAU)
Additional services/ provisions	AWS/On-Premises Hybrid cloud, Time-series DB, Time-series virtualization dashboard
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 30 Titania (UC 3) – Intelligent Toolbox for OHSE Data

Dataset reference and name	Toolbox for OHSE Data
Description	Bio Signals Analytics & Anomaly Detection
Source	Dig_IT IIoT platform
Other partners contributing	ICCS, CORE,TITANIA, SEI
Personal data	Yes
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	GDPR, Article 9 GDPR
Technology	N/A
Standards and Metadata	Neural network (Encoder – Decoder architecture) consuming multi-variate time-series from the sensor data that are smart garment.
Data sharing	LIBRA No open data will be generated directly
Data volume	N/A
Data type and format	JSON schema
Update frequency	N/A
Archiving and presentation	Dig_IT IIoT platform, Dashboard (LIBRA).

Dataset reference and name	Toolbox for OHSE Data
Responsibilities	Dig_IT IIoT platform
Resources	Dig_IT IIoT platform
Additional services/ provisions	AWS/On-Premises Hybrid cloud, Time-series DB, Time-series virtualization dashboard
Ethical aspects	Biometric data. Processing of biometric data for the purpose of uniquely identifying natural persons is prohibited. The user's heart rate information and identification information are sensitive and must be compliant with GDPR rules. Data integrity and security needs to be ensured.
Other	N/A

Table 31 Titania (UC 3) – Intelligent Toolbox for OHSE Data

Dataset reference and name	Toolbox for OHSE Data
Description	Air-Quality Smart Monitoring and Forecasting
Source	Dig_IT IIoT platform
Other partners contributing	ICCS, CORE, SINTEF, TITANIA, SEI
Personal data	No
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	N/A
Standards and Metadata	Neural network (LSTM and/or Transformer architectures) consuming multi-variate time-series sensor data
Data sharing	LIBRA No open data will be generated directly.
Data volume	N/A
Data type and format	JSON schema
Update frequency	N/A
Archiving and presentation	Dig_IT IIoT platform, Dashboard (LIBRA).
Responsibilities	Dig_IT IIoT platform
Resources	Dig_IT IIoT platform
Additional services/ provisions	AWS/On-Premises Hybrid cloud, Time-series DB, Time-series virtualization dashboard
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

3.2.1.4 UC4 – Hannukainen mine (open pit)

The Hannukainen Mining Oy is a subsidiary of Tapojarvi and it is based on the Municipality of Kolari. The Hannukainen project is planned to be on two different sites (Hannukainen mine and Ratuvaara Mill) that are connected with an overland conveyor. In particular Hannukainen Mining Oy is essentially engaged in four main activities: Scaling, Rock supporting, production drilling and loading.

TAPO operates in most underground mines in Finland (~220 mining machinery and vehicles). TAPO as mining operator offers the whole underground production chain: tunnel drilling, charging and

blasting, short hole bolting and meshing, concrete spraying, long hole bolting and drilling, loading and transportation of blasted rock, scaling, mounting, boulder hammering, watering, crushing, road maintenance. TAPO aims to advance their current asset monitoring (i.e., routes, vibration, consumption, emissions, operating time, temperature, pressure, load) and respective productivity and environmental performance. The most critical parts of their customer's value chain are the safety and reliability of transportation and loading of ore and waste rock. Respectively TAPO's most important assets are the loading and hauling equipment and the drilling rigs.

Table 32 Hannukainen (UC 4) – EHS & Sustainability data - Water quality, laboratory results dataset

Dataset reference and name	Water quality, laboratory results data
Description	Laboratory test results from samples that are taken regularly from the streams and rivers around the planned mine site
Source	Laboratory test results that are compiled in to excel
Other partners contributing	Sintef, Sintef Norlab, Laboratory
Personal data	No
License	All rights reserved to Hannukainen Mining. Majority of the data is publishable
Embargo periods	N/A
Legal framework(s)	N/A
Technology	TBD
Standards and Metadata	Laboratory test results from various samples are compiled in to one excel. The sampling and analysis is done by an accredited laboratory. The excel itself is not done according to any specific standard.
Data sharing	All rights reserved to Hannukainen Mining. Majority of the data is publishable. Please ask for publication permission from Hannukainen personnel
Data volume	< 10Mb
Data type and format	Excel-file
Update frequency	Updated three times a year when lab results are received from the laboratory
Archiving and presentation	The file and data is stored at the Hannukainen's server
Responsibilities	Environmental engineer's
Resources	The sampling and laboratory tests are done three times a year by an independent and certified laboratory. Personnel doing the sampling have also been certified.
Additional services and provisions	Data security done according to Tapojärvi and Hannukainen's data management policy.
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 33 Hannukainen (UC 4) – EHS & Sustainability data - Water quality, online measurement dataset

Dataset reference and name	Water quality, online measurements data
Description	Water quality data that is gathered on-line from the measuring stations
Source	On-line measurement stations feed the data to EHP's server, where it can be accessed
Other partners contributing	Sintef, Sintef Norlab, EHP Environment
Personal data	No

Dataset reference and name	Water quality, online measurements data
License	All rights reserved to Hannukainen Mining. Majority of the data is publishable, when permission is asked and the data is reviewed by Hannukainen personnel before publishing
Embargo periods	N/A
Legal framework(s)	N/A
Technology	TBD
Standards and Metadata	Online measurements (pH, turbidity, conductivity, water level, water temperature) are gathered in to a database at EHP's server. Date, place and time of measurement is also gathered. Also, data (battery status, errors, etc.) from the station itself is gathered.
Data sharing	All rights reserved to Hannukainen Mining. Majority of the data is publishable. Please ask for publication permission from Hannukainen personnel
Data volume	TBD
Data type and format	Database. Type of data base not known at the moment.
Update frequency	Can be changed according to measurement needs
Archiving and presentation	The file and data is stored at the EHP Environment's server. Can be accessed via internet service portal.
Responsibilities	Environmental engineers are responsible for the monitoring of the data and devices. Maintenance services to the stations are from EHP Environment.
Resources	The data quality is monitored by cross checking them to the lab results. This is done manually.
Additional services and provisions	Data security done according to EHP's, Tapojärvi's and Hannukainen data management policy.
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 34 Hannukainen (UC 4) – EHS & Sustainability data - Weather station dataset

Dataset reference and name	Weather station data
Description	Data from the weather station in Hannukainen mine site. Data is gathered online.
Source	On-line measurement stations feed the data to EHP's server, where it can be accessed.
Other partners contributing	Sintef, Sintef Norlab, EHP Environment
Personal data	No
License	All rights reserved to Hannukainen Mining. Majority of the data is publishable, when permission is asked and the data is reviewed by Hannukainen personnel.
Embargo periods	N/A
Legal framework(s)	N/A
Technology	TBD
Standards and Metadata	Online measurements (temperature, moisture, amount of rain (in the summer), wind speed and direction, air pressure) are gathered in to a database at EHP's server. Date, place and time of measurement is also gathered. Also, data (battery status, errors, etc) from the station itself is gathered.
Data sharing	All rights reserved to Hannukainen Mining. Majority of the data is publishable. Please ask for publication permission from Hannukainen personnel
Data volume	TBD
Data type and format	Data base. Type of database not known at the moment.
Update frequency	Can be changed according to measurement needs

Dataset reference and name	Weather station data
Archiving and presentation	The file and data is stored at the EHP Environment's server. Can be accessed via EHP's internet service portal.
Responsibilities	Environmental engineers are responsible for the monitoring of the data and devices. Maintenance services to the station are from EHP Environment.
Resources	TBD
Additional services and provisions	Data security done according to Tapojärvi and Hannukainen's data management policy.
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

3.2.1.5 UC5 – Sotkamo Silver mine (underground mine)

Hopeakaivos (Silver mine owned by Sotkamo Silver) is TAPO's asset-focused use case. Through Dig_IT, TAPO aspires to address their needs in predictive operation, predictive maintenance, optimal scheduling, operations management optimization, energy consumption reduction through new technologies and on-stream data analytics. Digitization of the Sotkamo Silver underground mine environment will permit to achieve the EHS methodologies to increase the workers safety, providing an information flow for the local community regarding the environmental compliance of their operations by integrating environmental compliance monitoring to facilitate communication with local stakeholders. These guidelines will be carried out with real time measurements of different parameters, such as temperature variations, rock dust, air quality, noise, vibration, etc., or with technologies to support human vision and hearing in charging and other skills in loading and transport or with application of technologies to increase hygiene at work.

Table 35 Sotkamo (UC 5) – EHS & Sustainability data - OHSE Dataset

Dataset reference and name	OHSE Data
Description	OHSE dataset – generic for mines. Sensors for surveying OHSE. Potential audience of interest: End users.
Source	Sensors
Other partners contributing	End user: Tapojärvi, Sotkamo
Personal data	No
Licence	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	Not identified yet.
Standards and Metadata	From D1.5: Sensors for online measurement of OSH-parameters: Noise Dust Diesel exhaust (NO ₂ and CO) Blasting gases (NO ₂ , CO and NH ₃) – stationary measurements

Dataset reference and name	OHSE Data
	<p>OSH- parameters: Flow charts of data flow is described in fig. 3-1, 3-2, 3-3 in D1.5</p> <p>Data will be created from scratch (on-line measurements and at-line measurements) and statistical treated data based on these.</p> <p>Metadata protocol will be decided by other partners in Dig_IT.</p>
Data sharing	Not decided yet. To be decided in cooperation with the end user how and where the data will be displayed at the end user. Dissemination of the data will be decided upon in cooperation with the end users. No open data will be generated directly.
Data volume	Depending on data structure and data flow in the project.
Data type and format	N/A
Update frequency	Suggested in the flowcharts.
Archiving and presentation	Treated by other partners.
Responsibilities	N/A
Resources	N/A
Additional services and provisions	Treated by other partners
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 36 Sotkamo (UC 5) – Situational awareness & biometrical data - Stress & Fatigue Monitoring Dataset

Dataset reference and name	Stress & Fatigue Monitoring Data
Description	Data from users' wristbands, which is used to determine the overall stress and fatigue levels users. Potential audience of interest: End users.
Source	Heart rate sensor – PPG Accelerometers/gyroscopes Skin Temperature GalvanicSkinResponse Sensor - GSR (“micro-Mho (μM)” or “micro-Siemens (μS)”)
Other partners contributing	End users: Sotkamo, ICCS
Personal data	Yes
License	Not to be disclosed outside of consortium without consent of user
Embargo periods	N/A
Legal framework(s)	GDPR, article 9 (2) GDPR
Technology	Wristband developed by ICCS, please refer to Deliverable 2.4.
Standards and Metadata	From D1.8: Examine collapse, muscle cramps, heart palpitation, carelessness, accidents, mental health problems, chronic health issues
Data sharing	Not decided yet. To be decided in cooperation with end user how and where the data will be displayed at the end user. Dissemination of the data will be decided in cooperation with the end users. No open data will be generated directly.
Data volume	Depending on data structure and data flow in the project.
Data type and format	Proposed: JSON

Dataset reference and name	Stress & Fatigue Monitoring Data
Update frequency	Not decided yet. It will be decided based on the smart garment capabilities/limitations (e.g., device battery drainage, processing requirements) and use case needs.
Archiving and presentation	Data will flow from the end devices to the ICCS platform that will be delivered in WP3 and detailed in D3.1 (Dig_IT IIoT platform architecture design). The accumulated data will be stored at the platform and exploited by other partners (e.g., LIBRA, CORE) for analytics services and visualization. The data preservation policy will be decided in collaboration with the end users (Mines) and use case requirements.
Responsibilities	All partners present on the data flow including data production, transmission, processing and consumption.
Resources	<ul style="list-style-type: none"> • Sensors that collect/measure the biometrics data. • Bluetooth communication module that interconnects the wristband to the smart garment. • Smart garment main processing unit that collects sensed data (format, prepare and send over WiFi). • WiFi connectivity to expose the data outside the mines area(s) towards the Dig_IT partners and platforms. <p>Please refer to D2.4 for more details regarding the smart garment components (wristband, headset, garment).</p> <p>Data maintenance will be decided by the end users and partners involved in data acquisition, processing and storage as will be detailed in WP3 and WP5.</p>
Additional services and provisions	N/A
Ethical aspects	Biometric data. Processing of biometric data for the purpose of uniquely identifying natural persons is prohibited. The user's heart rate information and identification information are sensitive and must be compliant with GDPR rules. Data integrity and security needs to be ensured.
Other	N/A

Table 37 Sotkamo (UC 5) – Situational awareness & biometrical data - Location Dataset

Dataset reference and name	Location Data
Description	Location of users in the mine
Source	Wearable device
Other partners contributing	End users: Sotkamo, ICCS
Personal data	Yes
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	GDPR
Technology	The technology that will be exploited for the localization service will be based on ultra-wideband (UWB) technology. Please refer to D2.4 for more details.
Standards and Metadata	Not yet available

Dataset reference and name	Location Data
Data sharing	Not decided yet. To be decided in cooperation with end user how and where the data will be displayed at the end user. Dissemination of the data will be decided in cooperation with the end users. No open data will be generated directly.
Data volume	Depending on data structure and data flow in the project.
Data type and format	Proposed: JSON file indicating x,y,z coordinates of the person within the mine area.
Update frequency	Not decided yet. It will be decided based on the smart garment capabilities/limitations (e.g., device battery drainage, processing requirements) and use case needs.
Archiving and presentation	Real time stream that does not require storage. ICCS generates the location info that is exploited by the respective task in WP5.
Responsibilities	ICCS and task partners in WP5.
Resources	<ul style="list-style-type: none"> • A network of UWB sensors placed within the mine. • Smart garment to transmit the calculated position based on the UWB metrics to the ICCS platform and task partners of WP5 <p>Please refer to D2.4 for more details.</p>
Additional services and provisions	N/A
Ethical aspects	Personal data. The user's location information is sensitive and must be compliant with GDPR rules. Data integrity and security needs to be ensured
Other	N/A

Table 38 Sotkamo (UC 5) – Situational awareness & biometrical data - Wearable OHSE Dataset

Dataset reference and name	OHSE Wearable Data
Description	OHSE dataset – generic for mines. Sensors for surveying OHSE. Potential audience of interest: End users.
Source	Ambient temperature sensor Humidity Sensor Air quality sensors (e.g., CO, CH ₄ , NH ₃ , NO _x , CO ₂) Noise level sensors
Other partners contributing	End user: Sotkamo, ICCS
Personal data	No
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	Detailed information for the sensor types used and means of connectivity among them are detailed in D2.4.
Standards and Metadata	Please refer to D1.5 and D2.4.
Data sharing	Not decided yet. To be decided in cooperation with end user how and where the data will be displayed at the end user. Dissemination of the data will be decided upon in cooperation with the end users. No open data will be generated directly.
Data volume	Depending on data structure and data flow in the project.

Dataset reference and name	OHSE Wearable Data
Data type and format	Proposed: JSON
Update frequency	Not decided yet. It will be decided based on the smart garment capabilities/limitations (e.g., device battery drainage, processing requirements) and use case needs.
Archiving and presentation	Data will flow from the end devices to the ICCS platform that will be delivered in WP3 and detailed in D3.1 (Dig_IT IIoT platform architecture design). The accumulated data will be stored at the platform and exploited by other partners (e.g., LIBRA, CORE) for analytics services and visualization. The data preservation policy will be decided in collaboration with the end users (Mines) and use case requirements.
Responsibilities	ICCS and task partners in WP5.
Resources	<ul style="list-style-type: none"> • OHSE sensors. • Smart garment main processing unit that collects sensed data (format, prepare and send over WiFi). • WiFi connectivity to expose the data outside the mines area(s) towards the Dig_IT partners and platforms. <p>Please refer to D2.4 for more details regarding the smart garment components (wristband, headset, garment).</p> <p>Data maintenance will be decided by the end users and partners involved in data acquisition, processing and storage as will be detailed in WP3 and WP5.</p>
Additional services and provisions	N/A
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 39 Sotkamo (UC 5) – Digital Twin Air Quality Data

Dataset reference and name	Digital twin air quality data (Sotkamo)
Description	Air quality properties (Pollutant's concentration) inside the quarry used by the final user for information, decision supporting and optimization tasks.
Source	Digital twins (virtual sensors) embedded into the DIG_IT platform.
Other partners contributing	Sinteff: Data acquisition. Schneider: Data aggregator. LIBRA: Decision support system. Sotkamo: Final user
Personal data	NO
License	N/A
Embargo periods	N/A
Legal framework(s)	N/A
Technology	Digital twin generated by using ITAINNOVA in house CFD model (OpenFOAM based) reduced order model library (Twinkle)
Standards and Metadata	Real time assessment of the air quality inside the mine is generated by the digital twin. Pollutant's concentration during working hours as well as pollutant's evacuation time after blasting.

	The data can be used to generate risk maps and alerts that support the final user for optimization and improvement of air quality during working hours. Information of pollutant's evacuation time after blasting can reduce the inactivity time.
Data sharing	Data is generated inside the DIG_IT platform and shown in the platform DSS, which is accessible by the final user.
Data volume	< 1Mb.
Data type and format	Image formatting and text or csv files.
Update frequency	Digital twin is based on steady state CFD simulations, representing the different operations inside the mine (blasting, loading...), so update frequency will be that one of the associated activity being done.
Archiving and presentation	Data is presented in the platform through the user interface (DSS) developed by LIBRA.
Responsibilities	N/A
Resources	N/A
Additional services and provisions	Other partners in charge.
Ethical aspects	No personal data. No ethical aspects identified.
Other	

Table 40 Sotkamo (UC 5) – Situational awareness & biometrical data - Smart Garment Command & Control Dataset

Dataset reference and name	Smart Garment Command & Control Data
Description	Data for controlling devices & commands from users.
Source	Smart earplug/headset
Other partners contributing	End user: Sotkamo, ICCS
Personal data	No
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	Please refer to D2.4.
Standards and Metadata	Not decided yet.
Data sharing	Not decided yet. To be decided in cooperation with end user how and where the data will be displayed at the end user. Dissemination of the data will be decided upon in cooperation with the end users. No open data will be generated directly.
Data volume	Keyword that is detected.
Data type and format	Proposed: JSON
Update frequency	Event driven (triggered by the smart garment user)
Archiving and presentation	Detected keyword will reach the ICCS platform and potential trigger alert back to the user. The data preservation policy will be decided in collaboration with the end users (Mines) and use case requirements.
Responsibilities	ICCS and end users.
Resources	Headset subsystem as described in D2.4
Additional services and provisions	N/A
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 41 Sotkamo (UC 5) – Assets data – Maintenance log Dataset

Dataset reference and name	Maintenance log Data
Description	Maintenance log data from Tapojärvi machines
Source	Maintenance database.
Other partners contributing	End user: Sotkamo, SUBTERRA
Personal data	N/A
License	N/A
Embargo periods	N/A
Legal framework(s)	Dig_IT NDA.
Technology	Research on APIs ongoing.
Standards and Metadata	Historical maintenance data logged by users as freeform text. Currently only historical data is accessed. Real time transfer of sensor data is being developed. Once an initial dataset is complete, metadata description of it will be added.
Data sharing	The data can be shared with the Dig_IT research partners. Currently the plans for sharing it are not finalised, as the dataset is not complete.
Data volume	Unknown data collection still being developed.
Data type and format	Maintenance data logs as Excel files. Include time, project, type, class, operation and machine usage hours data.
Update frequency	Unknown.
Archiving and presentation	The data will be initially archived and backed up on secure TAU hard drives, and later shared with other research partners through means chosen together. Data will not be archived beyond the duration of the project.
Responsibilities	The data will be managed by the TAU research group.
Resources	Tapojärvi IT department, Tapojärvi project partners, TAU research group.
Additional services and provisions	Uncertain at this time, as the collection of the dataset is still being developed.
Ethical aspects	N/A
Other	

Table 42 Sotkamo (UC 5) – PMA Data

Dataset reference and name	Predictive Maintenance Agents (PMA) predictions
Description	Future States and Failures in RT. Individual assets health (machine, tools and vehicles) with related historical data (ID: PMA_001) Predictive maintenance with regards to emissions (ID: PMA_002)
Source	Dig_IT IIoT platform, TAU
Other partners contributing	TAU, SOTKAMO, CORE, SEI
Personal data	No
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	N/A

Dataset reference and name	Predictive Maintenance Agents (PMA) predictions
Standards and Metadata	Neural network (Encoder – Decoder architecture) consuming multi-variate time-series from the sensor data that are placed on each asset. Neural network (LSTM and/or Transformer architectures) consuming multi-variate time-series sensor data placed on the asset
Data sharing	LIBRA No open data will be generated directly.
Data volume	N/A
Data type and format	JSON schema
Update frequency	N/A
Archiving and presentation	Dig_IT IIoT platform, Dashboard (LIBRA).
Responsibilities	Dig_IT IIoT platform
Resources	RT-DT (TAU), Dig_IT IIoT platform, Maintenance data (TAU)
Additional services/ provisions	AWS/On-Premises Hybrid cloud, Time-series DB, Time-series virtualization dashboard
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 43 Sotkamo (UC 5) – POS Data

Dataset reference and name	Predictive Operation System (POS) predictions
Description	Downtime prediction per asset (ID: POS_001) Maintenance cost per asset (ID: POS_002) Energy Consumption (ID: POS_003)
Source	Dig_IT IIoT platform, TAU
Other partners contributing	TAU, CORE, SOTKAMO, SEI
Personal data	No
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	REST API
Standards and Metadata	Neural network (LSTM and/or Transformer architectures) consuming multi-variate time-series sensor data placed on the asset
Data sharing	LIBRA No open data will be generated directly.
Data volume	N/A
Data type and format	JSON schema
Update frequency	N/A
Archiving and presentation	Dig_IT IIoT platform, Dashboard (LIBRA).
Responsibilities	Dig_IT IIoT platform
Resources	RT-DT (TAU), Dig_IT IIoT platform, Maintenance data (TAU)
Additional services/ provisions	AWS/On-Premises Hybrid cloud, Time-series DB, Time-series virtualization dashboard
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 44 Sotkamo (UC 5) – Intelligent Toolbox for OHSE Data

Dataset reference and name	Intelligent Toolbox for OHSE
Description	Bio Signals Analytics & Anomaly Detection
Source	Dig_IT IIoT platform
Other partners contributing	ICCS, CORE, SOTKAMO, SEI
Personal data	Yes
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	GDPR, Article 9 GDPR
Technology	N/A
Standards and Metadata	Neural network (Encoder – Decoder architecture) consuming multi-variate time-series from the sensor data that are smart garment.
Data sharing	LIBRA No open data will be generated directly
Data volume	N/A
Data type and format	JSON schema
Update frequency	N/A
Archiving and presentation	Dig_IT IIoT platform, Dashboard (LIBRA).
Responsibilities	Dig_IT IIoT platform
Resources	Dig_IT IIoT platform
Additional services/ provisions	AWS/On-Premises Hybrid cloud, Time-series DB, Time-series virtualization dashboard
Ethical aspects	Biometric data. Processing of biometric data for the purpose of uniquely identifying natural persons is prohibited. The user's heart rate information and identification information are sensitive and must be compliant with GDPR rules. Data integrity and security needs to be ensured.
Other	N/A

Table 45 Sotkamo (UC 5) – Intelligent Toolbox for OHSE Data

Dataset reference and name	Intelligent Toolbox for OHSE
Description	Air-Quality Smart Monitoring and Forecasting
Source	Dig_IT IIoT platform
Other partners contributing	ICCS, CORE, SINTEF, SEI, SOTKAMO
Personal data	No
License	Not to be disclosed outside of consortium without consent of end user
Embargo periods	N/A
Legal framework(s)	N/A
Technology	N/A
Standards and Metadata	Neural network (LSTM and/or Transformer architectures) consuming multi-variate time-series sensor data
Data sharing	LIBRA No open data will be generated directly.

Dataset reference and name	Intelligent Toolbox for OHSE
Data volume	N/A
Data type and format	JSON schema
Update frequency	N/A
Archiving and presentation	Dig_IT IIoT platform, Dashboard (LIBRA).
Responsibilities	Dig_IT IIoT platform
Resources	Dig_IT IIoT platform
Additional services/ provisions	AWS/On-Premises Hybrid cloud, Time-series DB, Time-series virtualization dashboard
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

3.2.2 Other datasets

In this subsection, we report the rest of the datasets the partners are handling. In particular, we report the datasets for the Life Cycle Assessment (LCA), public website data, aggregator dataset coming from assets and going to ICCS platform, alerting related data, environmental emission data written in the Blockchain, end-User Needs in Table 46, Table 47, Table 48, Table 49, Table 50, Table 51 respectively. By this we extend the analysis covering all research parts of the project that involves research data by providing detailed datasets for all beneficiaries of the projects.

Table 46 Data for the Life Cycle Assessment (LCA)

Dataset reference and name	Data for the Life Cycle Assessment (LCA) Task 5.2: Sustainability Assessment related to IIoTp
Description	Different sets of data will be collected for the LCA modelling procedure. For the needs of this task, STRATAGEM will first collect average historical data of the mine operation in all the life cycle stages, then those data will be compared against the collected data after the installation of the IIoTp platform, so as to observe how the new installations affect the mine operation in practice in terms of environmental impact. The second part of this study includes the collection of all the energy and resource consuming installations of the IIoTp platform in the mines. This dataset will assess the environmental impact of the platform. The data for the datasets are collected from the mines and ICCS.
Source	Collect data through meetings and emails. Methodological templates have been developed.
Other partners contributing	Partners involved in the IIoTp design, and the mines will contribute to the task by providing the required data.
Personal data	No
License	N/A – respect to confidential data
Embargo periods	N/A
Legal framework(s)	N/A
Technology	SimaPro software (for modelling the LCA) and the Ecoinvent library.
Standards and Metadata	The results of the LCA modelling procedure will reveal the environmental impact of the IIoTp systems in the use cases,

Dataset reference and name	Data for the Life Cycle Assessment (LCA) Task 5.2: Sustainability Assessment related to IIoTp
	<p>as well as the environmental impact of the mining processes before and after the installation of the IIoTp system (in the use cases).</p> <p>The quality of the results strongly depends on the quality of the data received from the relevant partners as well as the lack of available data in the software.</p> <p>As a first step to collect the data, one-to-one meetings took place between STRATAGEM and the mines to understand the data that we will need. More meetings took place to guide them through the methodological templates and finally for requesting more data or explanations.</p> <p>The collected data are the datasets. By using the software SimaPro, we will run the simulations to translate the datasets to environmental impact. Therefore, the environmental impact results will be the metadata. every time that there is a data change, the metadata are affected.</p>
Data sharing	The data can be shared with the consortium partners anytime in various formats.
Data volume	The datasets and metadata of the modelling procedures will be presented mainly in graphs and tables on the final deliverable. The number of the graphs cannot be estimated at this point but it will be manageable.
Data type and format	After collecting the data from the partners, the environmental impact for each case will be presented mainly in graphs. Word, Excel and LCA software files are the means which will be used to calculate, model and present the results of the study.
Update frequency	The average historical data, as well as new data will be collected and updated when and if necessary, but the frequency is still unknown.
Archiving and presentation	Storage and backups of the relevant materials in local servers (PCs, laptops etc.) as the first level of storage and backup. Additionally, data and files will be self - archive and preserved in a Google Drive folder that has been created for the purposed of the project.
Responsibilities	STRATAGEM's LCA analysisist is responsible for this task. SINTEF is also contributing to this task. All partners need to provide input to the task.
Resources	The resources needed to carry out the work are a laptop with the SimaPro software installed on it.
Additional services/ provisions	Nothing more.
Ethical aspects	No personal data. No ethical aspects identified.
Other	N/A

Table 47 Public website data

Dataset reference and name	Public website data
Description	The data include Up-to-date information about the project, the use cases, the milestones and relevant outcomes, project documents, and partners' press release.

Dataset reference and name	Public website data
	The audience of interest can any one of the targeted audiences mentioned in D8.3
Source	Meeting, survey, project documents, partners' press release
Other partners contributing	All partners
Personal data	None
License	None, contents are public but permission for using the data is required.
Embargo periods	Referring to Article 8.4.2 of Dig_IT Consortium Agreement
Legal framework(s)	Referring to Article 8.5.1 of Dig_IT Consortium Agreement
Technology	WordPress (last release), a free open-source content management system for websites development
Standards and Metadata	The public website data are dissemination data. The metadata used are: links to partners', EC portal and public media websites; links to project's social media channels.
Data sharing	Contents are public but permission for using the data is required.
Data volume	Up to 10 GB, currently about 1 GB
Data type and format	Text (WordPress format), image (png, jpg, svg), video (mp4), documents (pdf)
Update frequency	According to the project progress or dissemination activities
Archiving and presentation	The data are presented in formats allowed by the content management system and related plugins. The data storage, backup, availability and long-term preservation are guaranteed by the web hosting service provider. Maximum 10 GB of storage have been preserved.
Responsibilities	ROTECH
Resources	Resources for elaboration, maintenance and data security are provided by the web hosting service provider.
Additional services and provisions	None
Ethical aspects	No ethical or legal issues which have an impact on data sharing. No personal data are elaborated or maintained in this dataset.
Other	None

Table 48 Aggregator dataset coming from assets and going to ICCS platform

Dataset reference and name	Aggregator dataset coming from assets and going to ICCS platform
Description	The dataset come from other partners (TAU, SUBTERRA and SINTEF) and they are aggregate in the aggregator and then send via MQTT to ICCS platform and stored in a SQL database
Source	TAU: Mobile asset (trucks ecc) SUBTERRA: Data from drones SINTEF Data from sensors (NO2, Dust, ecc)

Dataset reference and name	Aggregator dataset coming from assets and going to ICCS platform
Other partners contributing	Titania, Marini Marmi and La Parrila
Personal data	NO
License	Ecostruxure Machine Scada Expert
Embargo periods	N/A
Legal framework(s)	N/A
Technology	MQTT, HTTP, SQL, OPC UA, IIS, .exe
Standards and Metadata	The data are collected from different fieldbus and protocol and the send via MQTT and stored in SQL database. The aggregator is stored in a industrial PC with Windows OS and will be accessible only from with specific credentials. It is also possible enable the E-sign if the End-user required it. Another security level is also possible: different credentials with different rights (read/write, only read ecc). The data is not related to variables that come from people but only from devices or mobile asset
Data sharing	MQTT, SQL, OPC UA, http
Data volume	Depending on data structure and data flow in the project
Data type and format	JSON and string. This is a more generic format that it is possible
Update frequency	Depends from the data (typically from 1 seconds to few hours)
Archiving and presentation	The aggregator has a graphic interface that consent to modify the data (with the proper credentials) and to visualize it. The data will be stored in a SQL database that is installed on the local industrial PC.
Responsibilities	ICCS platform. After the team responsible for data management and data curation is the IT team of end-user
Resources	The resources are basically an industrial PC with Ecostruxure Machine Scada Expert Software
Additional services and provisions	Ecostruxure Machine Scada Expert runtime, SQL server, MQTT broker
Ethical aspects	We are not collecting personal data but we are investigating about the ethical aspects
Other	N/A

Table 49 Alerting-related Data

Dataset reference and name	Alerting-related Data
Description	Data generated when an alert is triggered. Specifically, the timestamp, the Alert id and the metric value that triggered the alert are generated and stored in the data warehouse.

Dataset reference and name	Alerting-related Data
Source	All the alerts are generated from data consumed from the ICCS platform, in particular OHSE data, Geotechnical data, Stress & Fatigue monitoring data and Location data.
Other partners contributing	All partners providing data to the ICCS
Personal data	YES
License	Not to be disclosed outside of consortium without consent of user
Embargo periods	N/A
Legal framework(s)	GDPR, article 9 (2) GDPR
Technology	Depending on whether the alert is prompt notification or an urgent Alert (see D1.7), the alerts will be evaluated either in the DSS directly reading data from the ICCS data warehouse or through the Kafka broker in a streaming fashion, respectively. In both cases the resulting alerts, when triggered will be stored back in the ICCS data warehouse and/or Kafka broker.
Standards and Metadata	Each alert event will contain the following information: Timestamp, metric value that triggered the alert, Alert id, Alert description and triggering threshold
Data sharing	Data will be written in ICCS data warehouse and Kafka broker. From there data will either be displayed through the DSS to the appropriate end users (based on the RBAC system) and notifications will be sent to end users through various means agreed such as emails.
Data volume	We do not expect a large amount of data because alerts are not supposed to get triggered often.
Data type and format	In the data warehouse table, each row will represent an alert triggered that will consist of the metadata mentioned above. In a similar fashion in the Kafka broker a row will be transformed into a JSON record.
Update frequency	Non regular updates. A new record will be inserted whenever an alert is triggered.
Archiving and presentation	These data are stored in the ICCS platform and they are presented both in specific sections of the DSS and to the alert recipients through pre-specified means, e.g., email.
Responsibilities	LIBRA
Resources	The alerting system needs to consistently and promptly consume OHS, location, biometric data and EHS data both from the TimescaleDB and the Kafka broker when applies.
Additional services and provisions	Data will be visualized in the DSS and delivered through means such as emails.
Ethical aspects	Handling of alerts triggered from personal data, e.g., biometric or location data, will be compliant with GDPR rules. Data integrity and security needs to be ensured.
Other	N/A

Table 50 Environmental emission data written in the Blockchain

Dataset reference and name	Environmental emission data written in the Blockchain
Description	Dataset about the environmental impact written in the blockchain
Source	Industrial IoT platform (Aggregator) SEI
Other partners contributing	Use case (mine) partners and IIoT platform Aggregator
Personal data	NO
License	N/A
Embargo periods	N/A
Legal framework(s)	N/A - Will be the same applying to the IIoT platform
Technology	Kafka over the IOTA Tangle Ledger - API for the IIoT platform to write into the IOTA Tangle Ledger to be defined
Standards and Metadata	Digital Twin to be defined, and compliant with OPC UA communication architecture. W3C decentralized identities
Data sharing	APIs for end user access and dashboard visualization (TBD)
Data volume	N/A
Data type and format	JSON/ Text
Update frequency	Almost real-time (1-10')
Archiving and presentation	Data on the IOTA Tangle has to be considered temporary; long term storage will be discussed based on use cases (integration with IOTA Chronicles node - Scylla DB - to be defined)
Responsibilities	IOTA IIoT platform aggregator provider/use cases (mine) owners
Resources	Transparent to IOTA work; data will be collected first by the IIoT platform aggregator and then further exported to the IOTA Tangle (type and visibility to be controlled by data owner/IIoT platform aggregator)
Additional services and provisions	OPC UA Gateway from IIoT Platform to IOTA Tangle; end-users APIs from IOTA Tangle to visualization dashboard
Ethical aspects	No personal data involved
Other	IOTA will not generate or acquire direct data from sources; however, the IOTA Tangle ledger will be used to share data between Dig_IT stakeholders and the general public. Most of such data will be aggregated by the IIoT platform aggregator first then exposed to the IOTA infrastructure. Decision on the type data to be exposed is left to responsible stakeholders owning the respective data source. Control on data access is also managed by data owners. Policies are enforced accordingly by the IOTA infrastructure.

Table 51 End-User Needs dataset description

Dataset reference and name	End-User Needs data
Description	Dig_IT Mining and quarrying industry needs responses. Data collected via office form, from Dig_IT end users. Survey was

Dataset reference and name	End-User Needs data
	focused on participating mining industries and extracted information about their needs towards a more sustainable, digital, and human centered future. To achieve this, a set of questions was listed for every challenge related with project's objectives, technologies, and mining industry trends. Participants were requested to answer these questions based on their experience, data collected, operational barriers, business opportunities and investment plans. Data is gathered and stored in a 28KB .xls file in Dig_IT TEAMS/WP8 Market Outreach/Files/End-User needs file.
Source	Survey
Other partners contributing	TAPOJÄRVI OY, Titania AS, Subterra, Marini
Personal data	No personal data
License	Consensus statement
Embargo periods	N/A
Legal framework(s)	N/A
Technology	Online Office Forms
Standards and Metadata	New set of statistical data were created analysing and visualising end-user needs for the purpose of PEDR
Data sharing	Data were anonymised and shared only within consortium members
Data volume	28KB
Data type and format	Use of Excel format available online at https://forms.office.com/
Update frequency	No updates.
Archiving and presentation	No need for long term preservation. Data are archived in project's TEAMS repository and only used to extract end-user needs presented at 1st PEDR.
Responsibilities	CORE Innovation Management team
Resources	N/A
Additional services and provisions	N/A
Ethical aspects	No personal data. No ethics applied
Other	CORE during M1-M12 conducted a research with the consortium to identify end-user market needs. The purpose was to identify the needs of four mining industries and match them with Dig_IT innovations to design value propositions for PEDR. The online form is online at https://forms.office.com/r/QXEe04nyvm
Responsibilities	CORE Innovation Management team
Resources	N/A
Additional services and provisions	N/A
Ethical aspects	No personal data. No ethics applied
Other	CORE during M1-M12 conducted a research

4 Resources, Security and Ethics

4.1 Allocation of resources

In the Project we distinguish two types of information, the project management information, e.g., documents, meeting minutes, deliverables, templates, dissemination material, etc.; and the software developed, use case data and test data. Given the characteristics of each type of information, each of them is allocated in a different repository.

The Microsoft Teams and SharePoint repository, is the tool that is being used as a private area for preparing and sharing documents and knowledge among the members of the Project.

All the individual members of Dig_IT partners have access, as guests. WP leaders and use case Leaders are free to create specific working spaces to deal with their day-by-day activities.

This repository is also used for backups or for archiving of documents.

As mentioned before, it is a private repository, so it doesn't involve any cost to the project. It is managed by ITAINNOVA, and the costs derived from this activity are already considered in the efforts of the project.

Ro Technology designed and implemented also a public project website for dissemination scope following the rules indicated in paragraph 4.3 for the correct use of the social networks.

Ro Technology will develop a Dig_IT Toolbox, accessible through the project website, in order to share the source code, use cases data and test data. The Toolbox will have one public access point and it will only be used as a portal to share the source code, related public documentation and information.

Each partner is free to choose whether to share information and, eventually, what information and in what way. In addition, each partner is free to choose where to keep the source code and related documentation (for instance, Git clients, partner's repositories, etc.).

The Toolbox will be composed of a home page, containing a short preview of all components, and a dedicated page, where each partner can share information about a specific component. The latter can be a generic description, related documentation, instructions to download it, source code and so on.

4.2 Data security

Referring to previous sections, for sharing data the project members will use the following environments:

- ✓ Microsoft Teams and SharePoint repository for documentation related with the project;
- ✓ Project website for dissemination scope.

One of the most important concepts in security is that there is no completely secure system. It is simply not possible to prevent any current and future attack. If a new way to attack a system is found in the future, a system designed in the past can do nothing against it. Nevertheless, what a system can do is to have state-of-the-art security when made public and allow to be updated when new vulnerabilities are found before it affects its users. And this is what our solutions do.

4.2.1 Microsoft Teams

According to Microsoft: “Microsoft Teams is designed and developed in compliance with the Microsoft Trustworthy Computing Security Development Lifecycle (SDL), which is described at Microsoft Security Development Lifecycle (SDL). The first step in creating a more secure unified communications system was to design threat models and test each feature as it was designed. Multiple security-related improvements were built into the coding process and practices. Build-time tools detect buffer overruns and other potential security threats before the code is checked in to the final product. “

For enterprise customers, Microsoft stores the data from their 365 services, like Teams or SharePoint, in data centres nearest to the business location provided when the company create its tenant.

In case of tenants created with a billing address in any European Union country, the data centres are located in the European Union under European laws.

Also, Microsoft Teams corporate not only integrates Office365 in their solution, but also has redundancy in the backend, making the data accessible even if a given server fails.

Obviously, as the data is stored at a trusted third party, it has the required trusted certificates. Both solutions promise long term preservation and as the biggest players in each of their segments of the markets, they are trusted by millions.

4.2.2 Project Website

Project site is used as a showcase to promote the Project through its objectives, its results and the whole Consortiums activities. It also shows the Project publications, public reports and other Project related material.

The security measures applied in the Project website are not very strong, due to the purpose of the website, the absence of confidential information and documents, as well as the absence of inputs sections which request sensible data. Therefore, a suitable level of security is implemented through a couple of security features: TLS and SiteScanner.

TLS, which stands for Transport Layer Security, is a well-known cryptographic protocol used to provide Privacy and data Integrity in several applications, even if it is mostly known for HTTPS securing. In the Project website a Let's Encrypt certificate is used, which encrypts all traffic passing through the website and ensures that sensitive information such as login credentials and credit card details remain protected. This type of certificate is provided by the Let's Encrypt Certification Authority, a non-profit CA which provide TLS certificates to 260 million websites. The certificate has a validity period of 90 days and the request / renewal is handled automatically by the SLL manager provided by the hosting service.

SiteScanner is a security system that detects and warns about the presence of Malware. The system allows to do on-demand scans of the website and scans daily for blacklisted domains and malware. The malware database is constantly updated, allowing SiteScanner to detect even the latest threats. Furthermore, weekly it sends a report about the daily scans to the user.

In addition, a daily backup is provided by the hosting service, which kept the data for 30 days. The backup feature helps to ensure system Availability.

The website contains a section dedicated to Privacy Policy, describing how we collect, store, process, and utilize information provided by the users. This Policy covers our treatment of information that is gathered by this website and applies to information collected by third party websites or services, or data collected in the context of employment. Furthermore, cookies are used to remember certain information, such as user preferences, the type of device used and other information, ensuring a more efficient operation of the site, enriching the visitor experience and providing information to the site owners.

4.3 Ethical aspects

Data sharing and the Open Access implementation to the scientific research process is clearly enhancing scientific progress, while also benefiting transparency and reproducibility. However, making specific types of data, for example personal data, public and accessible will challenge the aforementioned advantages. Several discussions have taken place on this respect, however a general consensus in the scientific community has not yet been reached.

This paragraph describes what impact the data in the project will have on ethical aspects, specifically when these data are related to personal information.

The partners involved in the project may be aware that neither during or after the project, any of the data shared/stored will have an impact on ethical, gender, or other personal circumstances.

To this regard, the partners must follow a set of practices while involved in data treatment based on GDPR guidelines:

- Ensuring that the citizens' rights on their own data are protected.
- Confidentiality and privacy must be managed so that shared datasets are fully and suitably anonymised or pseudonymised and people cannot be identified from shared data.
- Adequate, suitable, prior, informed, written consent from people whom data are being obtained from must be validated, keeping clear the objective and purpose of the data to be obtained from them.
- Analyse the potential impact and implications on further research, in a way that further results obtained from shared data (never personal data, anonymized or not) can be validated and trustworthy.

As a rule of thumb, partners and participants shall follow the statement:

“All participants in the project will conform to GDPR and the current legislation and regulations in the countries where the research will be carried out”

Partners must ensure that the ethical and societal aspects are incorporated into the design process from the beginning. They must also allow security and validation of operational efficiency and transparency, possible interpretative distortions, liability for possible damage, patentability, anonymisation of user data and privacy in compliance with the guidelines developed by the European Commission Expert Panel [3] that has developed and continues to develop guidelines for the design of reliable artificial intelligence systems, respecting the centrality of the human being.

Additional information on each of the individual Use Cases can be found below. This information accounts for any sensitive data that may be involved either during research, development, or storage of data for further purposes, and is only available for those UCs that deal with data that may have any ethical impact (UC1, UC3, UC5).

In particular two data sets concern with ethical aspects:

- **STRESS & FATIGUE MONITORING DATA** (They are biometric data. Processing of biometric data for the purpose of uniquely identifying natural persons is out of project scope and then prohibited. Anyway, the user's heart rate information and identification information are sensitive and must be compliant with GDPR rules. Data integrity and security needs have to be ensured)
- **LOCATION DATA** (The user's location information is sensitive and must be compliant with GDPR rules. Data integrity and security needs have to be ensured)

The compliance with GDPR rules, data integrity and security needs will be assured adopting the following principles:

- ✓ The data will be anonymized and associated only with the device worn on the participant's body. Anyway, the device will record position at all times within the site, so that it is possible to know when and where the participants are moving. This may be seen as an infringement of participants' rights, both legally and ethically, but it should be clear that the sole purpose of the pilot is to improve the site workers' own safety, not their control within the site.
- ✓ UC1, UC3 and UC5 partners will ensure compliance of the UE 679/2016 of the European Parliament and of the Council which entered into force from May 2018 on the General Data Protection Regulation (GDPR) of individuals with regard to the processing of personal data and on the free movement of such data.
- ✓ Collected data at all stages will be treated with respect to personal anonymity. Only, when necessary, participants' personal data will be legally obtained after an informed consent. Data will be securely accessed, and privacy protection measures will be undertaken proportionally to the risks involved and the sensitivity of the data, such as password protection, encryption in all transmissions, etc. Identification data will be encrypted and strictly separated from sensitive data such as health data.

- ✓ Users' personal data will be collected for processing, and undergo such processing, only if they are adequate, relevant, and not excessive in relation to the scope and the specified, explicit and legitimate purposes for which they will be obtained. Also, personal data subjected to processing will not be used for purposes incompatible with those for which they will be collected.
- ✓ Further processing of the data for historical, statistical or scientific purposes shall not be considered incompatible. In addition, personal data will be erased when they will have ceased to be necessary or relevant for the purpose for which they will be obtained or recorded. They will not be kept in a form which permits identification of the data subject for longer than necessary for the purposes for which they will be obtained or recorded. On a regular basis, the procedure will be determined by which, exceptionally, it will be decided to keep the entire set of particular data, in accordance with the specific legislation, because of their historical, statistical or scientific value.
- ✓ The researchers will ensure that outcomes will be reported and will not contravene the right to privacy and data protection. They also will carefully evaluate and report the personal privacy implications of the intended use or potential use of the research outcomes.

Ethics matters in dissemination activities and use of the Socials means

The following precautions will be adopted in the use of social networks:

Creation of a project site and updating of its web pages with news, reports, images, publications, useful links to other initiatives.

ETHICS ACTIONS ADOPTED:

- Information on the policy adopted for navigation within the site, on the use of "cookies" present on the site (and their distinctions), for surveys for statistical purposes on the number of visitors and their area of geographical origin;
- Possibility of subscribing to the project newsletter and registering for project seminars or conferences by filling in a form that must contain the note on respect for privacy;
- Voluntary submission of personal data by the user to contact the project coordinator or other contacts listed on the site.

Drafting of a periodic newsletter.

ETHICS ACTIONS ADOPTED:

- Preparation of the information note to be circulated to recipients to notify them that they receive the newsletter following a registration;
- Insertion, in the sending message, of the possibility of requesting the cancellation of one's name from the list of recipients.

Organization of project dissemination events (seminars, conferences, information days with stakeholders, training and educational initiatives).

ETHICS ACTIONS ADOPTED:

- Preparation of the information to be issued to those registered for the events
- Consent to the processing of data in the event that a photo or video service is planned
- See Privacy Authority: Guidelines on the use of cookies and other tracking tools

Opening a project account on social networks.

ETHICS ACTIONS ADOPTED:

- In order to disseminate the photographs taken during an event, the interviews made in video or images that contain recognizable people, it is necessary to have previously acquired, during that same event, the authorization to use them for exclusive purposes of the project communication, specifying the methods adopted.

5 Opinions from the ethics board regarding this deliverable

The members of the project ethics board, Isabel Marco and Ignacio Giménez, gave a favourable opinion on this deliverable.

6 Conclusion

This version of the Data Management Plan provides the collection of the data managed during the project. Each partner reported how they are handling the data during and after the period of the project, what type of data they are treating, what methodology or standard they are using, and whether they are sharing or preserving the data during and after the timespan of the project. Moreover, partners have described the data collection and access procedures and the Data Management Plan for the distributed ledger technology. Section 4 contains details about the description of the resources for data management and peculiar aspects relating to data security and ethics. Section 5 contains the opinion/approval of the Dig_IT ethics board with regards to the Data Management Plan.

7 References

- | Ref. | Title |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [1] | Guidelines on FAIR Data Management in Horizon 2020 V3.0 26 July 2016. |
| [2] | Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Bourne, P.E. (2016). The FAIR Guiding Principles for scientific data management and stewardship. <i>Scientific Data</i> , 3, 160018, doi: 10.2038/sdata.2016.18. |
| [3] | https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52011DC0882&from=EN . |
| [4] | H2020 Programme, Guidelines on FAIR Data Management in Horizon 2020 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf . |
| [5] | Distributed, C., & Union, E. (n.d.). Distributed Ledger Technologies and Data Protection in the European Union. 1–57. |