DIGIECOQUARRY
INNOVATIVE DIGITAL SUSTAINABLE AGGREGATES SYSTEMS

Horizon 2020 research and innovation programme (Nº 101003750)

Deliverable D10.4.

Data Management Plan (DMP)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 101003750

Deliverable report

Deliverable No.	D10.4	Work Package No.	WP10	Task/s No.	Task 10.4
Work Package Title		Project Management			
Linked Task/s Title		Data Management Plan (DMP) and Gender Action Plan and monitoring			
Status		Draft	(Draft/Draft Final/Final)		
Dissemination level		PU	(PU-Public, PP, RE-Restricted, CO-Confidential)		
Due date deliverable		2021-11-30	Submission date		2023-03-29
Deliverable version		DIGIECOQUARRY_D10.4_AKKA_V2_20211125__ABAUT_ANEFA_ITK_SIGMA_APP.docx			

Document Contributors

Deliverable responsible	AKKA
Contributors	Organisation
Sadeq Zougari	AKKA
Diego Ignacio Laza Martin	ABAUT
Maria Loeck	ANEFA
Jan Spaeth	ITK
Pierre Plaza	SIGMA
Pablo Gomez Martin	APP
Reviewers	Organisation
ALL	ALL

Document History

Version	Date	Comment
1.0 Draft1	2021-11-03	First version for WPL, PC revision and for Peer review
1.1	2021-11-25	Final version
2.0	2023-03-29	Final version with editorial corrections included

Disclaimer

Table of contents

List of Abbreviations

Abbreviation	Description
AI	Artificial intelligence
API	Application Programming Interface
BIM	Building Information Management
CA	Consortium Agreement
CAN	Controller Area Network
CFS	Certificate on the Financial Statements
CSV	Comma-separated values, open text format
DEQ	DIGIECOQUARRY
DGPS	Differential Global Positioning System
DMP	Data management plan
DoA	Description of Action
DPIA	Data protection impact assessments
DPO	Data Protection Officer
DWH	Data warehouse
DXF	Drawing Interchange Format, or Drawing Exchange Format
EB	Exploitation Board
EC	European Commission
FAIR	Findable, Accessible, Interoperable and Re-usable
GA	Grant Agreement
GDPR	General Data Protection Regulation
H&S	Health and Safety
IFC	Industry Foundation Classes
IoT	Internet Of Things
IP (R)	Intellectual property (Rights)
IQS	Innovative Quarrying System
IWMS	Integrated Workplace Management System
JPEG	Joint Photographic Experts Group, recording format for a compressed digital representation of a still image
JSON	JavaScript Object Notation, lightweight format for storing and transporting data
KPI	Key Performance Indicator
KTA	Key Technology Area

Abbreviation	Description
OEM	Original Equipment Manufacturer
PDF	Portable document format, text file format
POPD	Protection of Personal Data
SQL	Structured Query Language used by databases
SW	Software
TXT	file extension for a Text file
WLAN	Wireless Local Area Network
WP	Work Package
XML	eXtensible Markup Language
ZIP	file compression format

1 Executive Summary

This document constitutes the first version of the Data Management Plan of the DIGIECOQUARRY project.

The Data Management Plan is an important document for efficient data management. It describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project and the process of making scientific publications and their research data findable, accessible, interoperable and re-usable (FAIR) according to article 29.2.

Consortium members will be provided with instructions, processes and templates to carry out data management in accordance with current European regulations. All this without prejudice to the most restrictive requirements that may be applicable in their national legislations.

It sets out the Data Management Plan that the DEQ project must comply with, explaining how the data is relevant and limited to the purposes of the research project. It provides information on identified challenges and summarises fundamental requirements to deal with privacy, data protection and other related issues in the project.

DIGIECOQUARRY (DEQ) Project includes research with humans. Amongst others, planned activities include surveys, workshops, participatory processes to assess acceptance of AI-based Health & Safety solutions and to collect feedback of different communities and local representatives. The collected data may include age, gender, data on health and safety as well as other social parameters. In addition, possibly confidential company data on the workplace's design, work processes and other working conditions are to be recorded. Requirements for data protection, confidentiality, risks assessment will be part of this deliverable.

This deliverable corresponds to Task 10.4 of Work Package (WP) 10 'Data Management Plan (DMP) and Gender Action Plan and monitoring'. The DMP will be updated at M47 (April 2025) according to the progress of the project.

2 Introduction

2.1 Objectives of the project

DIGIECOQUARRY is a Horizon 2020 project aiming to design, develop and validate in 5 pilot environments an Innovative Quarrying System (IQS) comprising sensors, processes, tools and methods for data capture, processing and sharing to provide integrated, digitalised, automatic and real-time process control for aggregates quarries.

The DIGIECOQUARRY consortium will combine the latest researched and advanced technologies applied to quarry operation together with the integration of selected innovative digital solutions to boost the capacity of the aggregates industry, to enhance Health & Safety conditions for workers, to improve the Process and Efficiency of the aggregates extractive sites, to maximise Sustainability and Resource Efficiency in the quarry operations and to foster Social Acceptance.

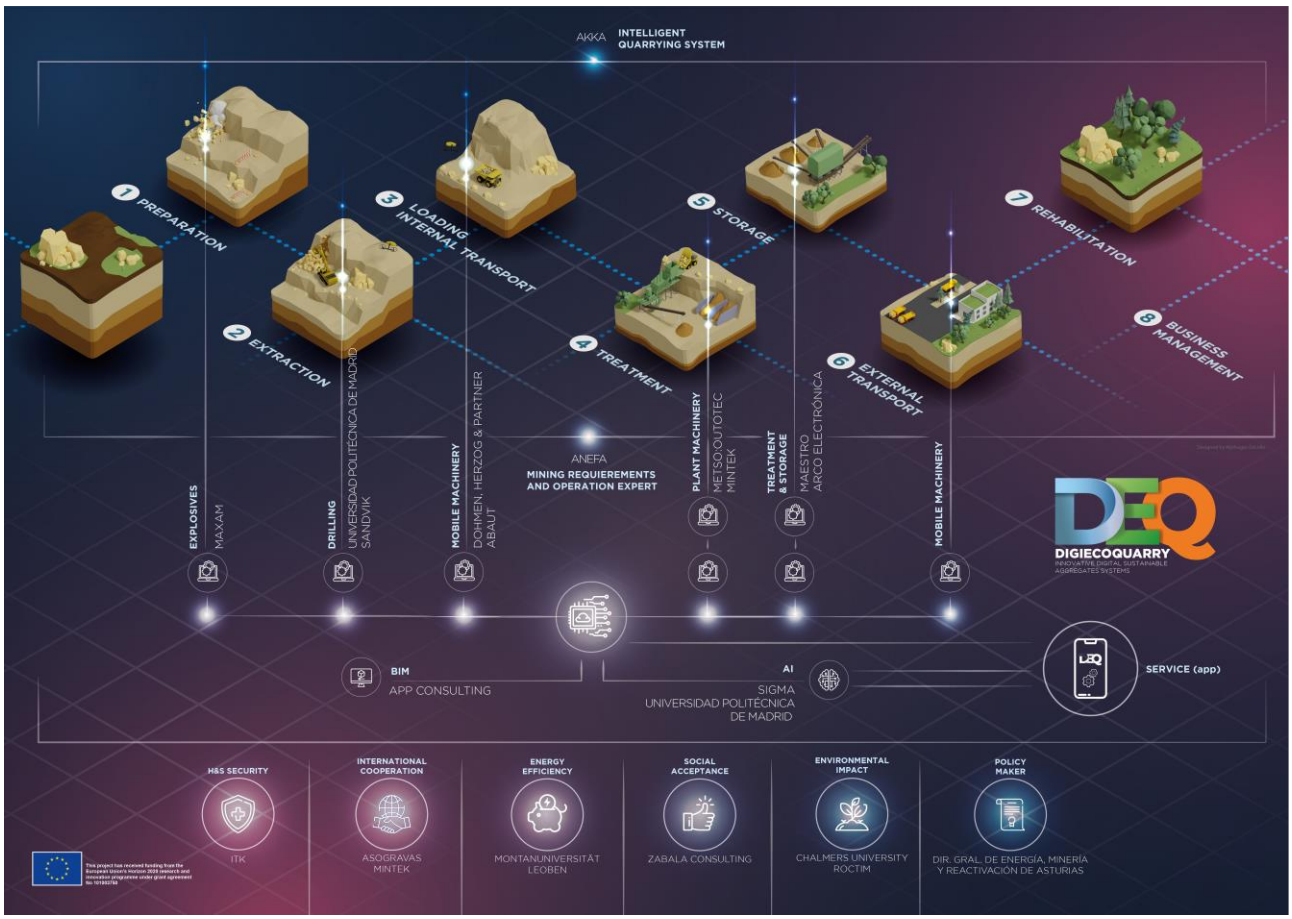

Figure 1. DIGIECOQUARRY concept.

2.2 Purpose of the document

This deliverable materializes the first version of the Data Management Plan elaborated by DIGIECOQUARRY. The purpose of this document is to provide an overview of the dataset types present in the project and to collect the main data management policy adopted by the Consortium.

The data management plan defines how data in general and research data in particular, will be handled during the research project, and will make suggestions for the after-project time. It describes what data will be collected, processed or generated, which methodologies and standards shall be followed during the collection process, whether and how these data shall be shared and/or made open in the context of "Open access to scientific publications and underlying research data" defined by article 29.2.

A first list of potential datasets that the project will make available is described.

This deliverable will be updated accordingly to the evolution of the project. Two versions are planned:

- The initial version D10.4 (M06, i.e. November 2021)
- The final version D10.11 (M47, i.e. April 2025)

2.3 Intended audience

The dissemination level of D10.4 is 'public' (PU) and available to members of the consortium, the Commission (EC) services and those external to the project.

This document is primarily intended to serve as an internal guideline and reference for all DIGIECOQUARRY beneficiaries, especially the governance bodies such as the General Assembly, the Project Management Board, the Project Technical Committee, and the International Advisory Board.

3 Data in DIGIECOQUARRY : an overview

3.1 Data management methodology in DIGIECOQUARRY

Data Management Plan (DMP), is a brief plan to define what data sets the project will generate or process, whether and how these data will be made accessible, and how they will be curated, stored and preserved. The DMP should also provide information on the measures taken to safeguard and protect sensitive data. Some personal information - such as a person's origin, political opinion, religious beliefs, health, trade union membership, or sexual orientation - is classified as sensitive personal data.

This document will help beneficiaries make their research data findable, accessible, interoperable and reusable (FAIR) as described in section 6.

The data generated will have different formats [e.g., Excel Tables, JPEG, CSV, DXF, etc.] according to the type and use. The project DigiEcoQuarry intends to generate the information needed for the Quarry of the future or Quarry 4.0. For doing this, the interoperability between data source and the different tools and applications should be considered as a must.

The size of the data, including the specific formats and the interdependency – connections between the sources – users – applications is still under definition.

The data collected, in any case, should be useful for one of the parties involved in the project not only for collecting, but for further quarrying and digitalization purposes.

The data produced as raw-data due to intellectual property (IP) rights, would only be shared under some specific circumstances and for specific applications. The process data and the results of the expert systems [solutions of the technology providers] will be accessible and will be shared with the rest of the partners.

The naming, re-use, versions, metadata and standards to use are under discussions for the project. Also, the methods, software, sharing data tools as well as relevant software, associated documentation and repository will be defined in further steps.

The data security and provisions will also be considered and studied to have a safe storage, handling and access to the data management. Also, the resources needed for the project and the responsibilities of the data management will be defined in future steps.

The following figure shows all the requirements, constraints, regulations, guidelines that were analysis for the definition of the DIGIECOQUARRY data management methodology. The DIGIECOQYARRY data collection process and data management is built upon requirements and methodologies coming from several processes defined in WP1-5 or guidance rules (Open Access, GDPR, ETHICS).

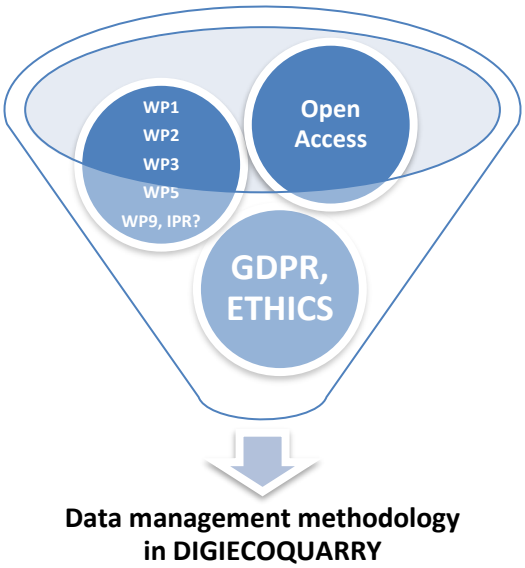

Figure 2: Data management methodology in DIGIECOQUARRY

3.2 Dataset categories

According to the project concept highlighted in Figure 1, this section describes project dataset categories.

Table 1. DigiEcoQuarry processes datasets general explanation and examples

Data category	General explanation	Data types examples*
Site Preparation	The information about the terrain's physical characteristics exploitation, but also about enabling roads in order for the mobile equipment to access efficiently all important points in the quarry.	3D Map, Images and videos, Twin model, quarry map.
Extraction	It includes all the information of the processes carried out during the extraction: Blast information, Explosive information, fuel consumption, water consumption, Quality of the material obtained.	Positioning (x,y,z) for drilling and explosives to be used, king and weight of explosives, litres of Diesel, Material Quality measures: Poor, medium, high, etc..
Loading & Internal Transportation	The information required here is relating to the process of loading mobile equipment from the extraction points to the treatment plant.	Distances, Weight, cycle times, workers information, dumping position, Hauling speed.
Treatment	The data processed here will consist of the production characteristics, the equipment performance statistics, the parameters used to control the production, lists of all consumables used in the process.	Tons of material, processing speed, Capacity of the treatment (t/h), Energy Consumption (Kw), Process configuration,
Storage	This will depend on the storage facilities if available on a Quarry. Normally storage in the case of a quarry relates to the time spent on	Types of material stored, weight of the materials in storage, area used, reaming capacity, etc…

Data category	General explanation	Data types examples*
	loading and transporting the materials produced at the treatment plant to the customers	
External Transport	In this data category, the data pertaining to the materials and external transportation facts including speed, distances, transporter facts, customer orders and receipts.	Material weight transported, monthly distribution, material type, etc…
Rehabilitation	The information on the backfilling of the quarry exploitation grounds is here handled.	Types of material used, Backfilling volume, videos of the areas of the quarry that are being rehabilitated, the kind of operations to recover the environment, etc…
Business Management	This category will be using data from all processes in a quarry basically that will help the quarry owners and managers to get via helpful dashboards daily and historical information on how the quarry is performing in general: the production, the processes, the workers, the market, its financial performance, etc..	Material weight, costs, consumables, workers, finances, KPIs, etc …
Equipment	A very significant amount of information will be available thanks to the sensors that are now installed in most of the quarry equipment: for example drilling machines, treatment plants, trucks, cranes, etc....	The data types & formats are specified by each equipment provider.
Safety	In order to access the safety a lot of data from the other data categories can be used (e.g. equipment, sensor data). When there is no data available, we expect to install different types of sensors depending on the use case.	Working Conditions of the System (running, running with limitations, on hold).
Environment	To gain relevant understanding of the environmental impact at the quarry, data from other categories will be used and processed to reflect environmental performance. Any missing data points will be attained through manual data input or other data collection methods dependant on the individual requirements.	Total Climate Change per tonne product, total Climate Change per month/year, Acidification per tonne etc.

*: Only preliminary data types are mentioned. A complete description will be provided in D3.1 - List and characterisation of key data inputs due in M12.

3.2.1 Data of the project that may be impacted by the regulations

This section provides for each KTAs a first preliminary analysis of the usage of personal data. This analysis will be completed according to regulation described in section 8.

- **Devices for automation of treatment plants and storage facilities**

Not applicable. There is no data in this area that may be subject to regulation issues.

- **Monitoring sensors and analysing tools for Mobile Machinery**

Some of the data involved under the regulation of the GDPR is the picture data. For tackling this issue, Abaut have already developed a sensor camera system able to generate, detect and blur the picture under the GDPR compliance. The picture data, in this case, is completely anonymous.

- Mobile equipment & quarry geological deposit digitalisation & real-time modelling

Not applicable. There is no data in this area that may be subject to regulation issues.

- BIM Management & Connection

Any personal data will not be involved. In case of any personal data will be involved, it will be compliant to GDPR rules.

- Artificial Intelligence (AI) algorithms

In the case of the services/applications that will be developed using artificial intelligence models, all data will be anonymised in order to be compliant with GDPR, such as:

  - Pilot workers personal Information (names, personal identification data, etc.)
  - Images or videos that may include persons (by blurring faces)
  - Providers/external workers personal information

Moreover, the amount of personal information will be the strictly minimal, needed to perform the analysis required by the models. In summary the principles described in section 8.1.3 will be ensured.

- Business management tool

All data fields corresponding to personal data can be affected regulations (in this case, for the Business management tool). Examples of personal data, but not only, that are required to develop the Business management tool are employee's data, customers data, partner companies or supplier data ... This type of personal data will follow the points and procedures described in *section Data Protection Regulation and Ethical Aspects* (and, if necessary, the deliverable "*D11.1 Ethics POPD Requirement*").

Some data that is processed in the Business management tool may not be subjected to regulation. Some of this are still extremely sensitive data for the company (e.g., billing details data). This implies that security standards and protection measures have to be taken into account to prevent unauthorized access to data or information leak.

- For the safety tool currently (November 2021) under development no personal data will be processed, stored or forwarded.

3.3 Data Collection and generation

The data generation and collection could come from any partners of the present project. The main application of these data is to be analysed by the expert systems. These systems are software's able to read and process the data generated for creating the process data needed by other expert systems of applications. According to the source, the data can be manually collected or generated by the technological infrastructure of the project.

3.3.1.1 Manually Collected Data

The manual collection will take at the Pilot Sites [Quarries] when an automated system is not/could not be integrated [e.g., Report of the machinery maintenance, consumption of fuel at the gas station, etc.]. This data

should be digitized in a defined format [e.g., Excel Table, Report, etc.] that can be integrated in the data lake and/or in the IQS for being used by the expert systems and by the rest of the partners.

The manually data collection should include all the relevant data needed for the interpretation and analysis by the rest of the partners within the consortium. This data should be shared in the Data Lake according to the principles ''Need-to-know'' to guarantee the data protection and the no-disclosure of confidential information.

The data, format and dependency of the data collected manually will be discussed in further steps of the project [e.g., maintenance plan, sales – sales forecast, etc.]

Pilot Sites :

Each expert system requires some initial data that should be provided by the Pilot Sites in order to start to create the data flow of the project. These different data sets can be:

- Machinery Lists and machinery maintenance plans
- Images & Maps from the sites [e.g., drone flights, DXF – topographic data, etc.]
- Consumptions [e.g., at the fuel station, electrical energy consumption at the transformation point of the site, etc.]
- Different kind forecast [e.g., sales, production, etc.]

This initial data set will set the initial correlation between data collected [Pilot] and application/usage [Expert System] of this data source [e.g., Machine 1 → energy consumption of Machine 1]

3.3.1.2 Data generated via the technological infrastructure of the project

The data generated via the technological infrastructure and digital tools is done by the different partners through their expert systems in collaboration with the pilot sites. The generation, handling, presentation and interpretation of the data will be done between the partners involved on these activities [e.g., Pilot Site – Technology Providers – University] in order to extract all the useful and meaningful information. This information will be correlated with the processes and KPI's to report. The definition of data requirements and characteristics, their formats and the dependency links with the data generated automatically will be defined in WP3-" Identification of key data inputs from cyber & physical quarrying system and development of sensors, automation and process control".

3.3.1.3 Data types and formats

Due to the different technologies applied, DigiEcoQuarry will create and modified at least this data formats according to the sources

- Processed data resulting from the Expert Systems and Pilot Sites
  - Numerical
  - Alphanumerical
  - Images & Maps
  - Dates

- Possible Format of the files
  - CSV, JSON
  - Excel (also in open data format)
  - JPEG

- TIFF
- Word (also in open data format)
- PDF
- Vector (point, line (or arc), and polygon data)
- Raster (grid data)
- IFC (https://en.wikipedia.org/wiki/Industry_Foundation_Classes)

- To be accessible to
  - Quarry Sites for data visualization, integration and further processes
  - Expert systems [other applications] for further processing

3.3.2 Sharepoint of the project (Teams)

During this project early stage, DigiEcoQuarry is working with SharePoint for sharing different files and documents and for schedule meetings according to the different working packages [WP] and tasks [T].

The layout of SharePoint goes as follows.

The main accessing application could be Teams or SharePoint. Through Teams we can observe the different layout in the following picture

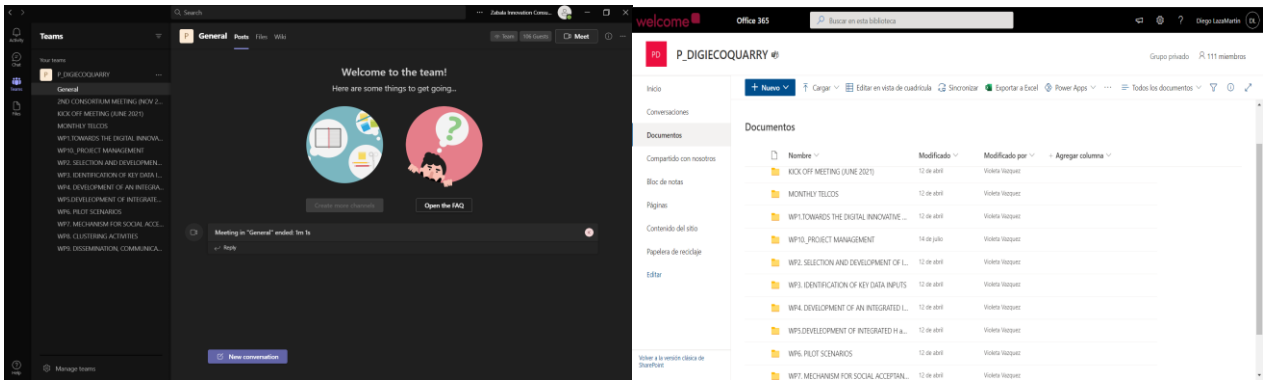

Figure 3: Teams data sharing platform and organization in DigiEcoQuarry

The project is using Teams as daily communication channel and SharePoint as file repository which helps for finding and sharing the different documents.

3.4 Data flows

The data flows that will be implemented in the Architecture of the IQS are represented in the figure below:

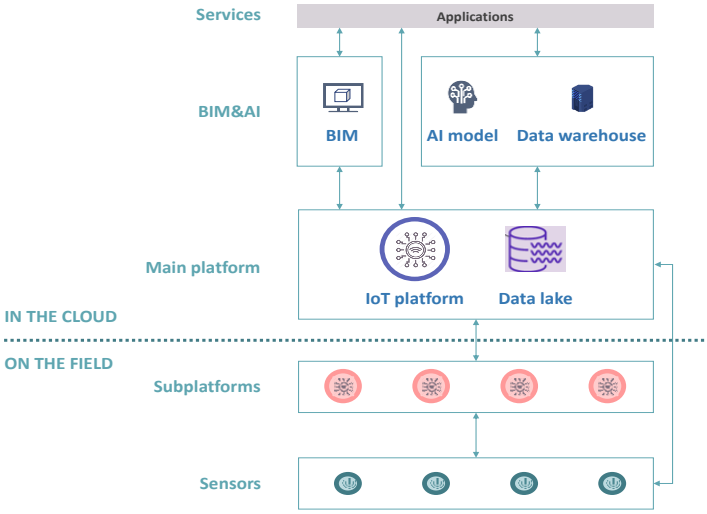

Figure 4: Data Architecture

Sensors are being and will be deployed along all the processes and the respective equipment that are used within the quarry production cycles. Sensors will send their data to their connected peers that are called sub-platforms in the Figure 4. The latter involve the communication needed and the data acquisition processes to further classify the data in a local server. These platforms may be accessed directly to show data on a per device, manufacturer level.

Every sub-platform will then transmit the needed optimised and pre-processed data to the Main Platform of the IQS, that will be implemented in the form of a Data lake, able to store and organize all the incoming data in many formats. In some cases, depending on the sensors deployed in the field, an IoT platform may be needed in the cloud to obtain the data in a similar way as for the sub-platforms already explained. The sub-platfoms are subsequently named in the chapter as “Expert Systems”.

The IoT platform will also send its refined data to the Data lake. In this way, we have a general repository of all data pertaining to a single quarry. In the figure, data coming from manual input and databases in the quarries is not displayed. The Data lake data will be accessible from the BIM and postprocessed to be stored in a structured fashion after applying data preparation cycles and Machine learning algorithms (AI Models) focused on generating the information needed by the AI services, Business management tools and dashboards at the upper level. The higher level structure of the Datawarehouse will be divided on separated data repositories for each quarry Data lake (Pilot sites in the project).

3.5 DIGIECOQUARRY components involved in digitalisation

In order to understand the data architecture at a high level we will use the following high-level schema displayed in the figure below. Data will be coming from live readings in real time or at discrete intervals form sensors, from data loggers installed in different equipment (mobile mainly) that are used to store data temporarily due to the poor connectivity that may be experienced in a quarry.

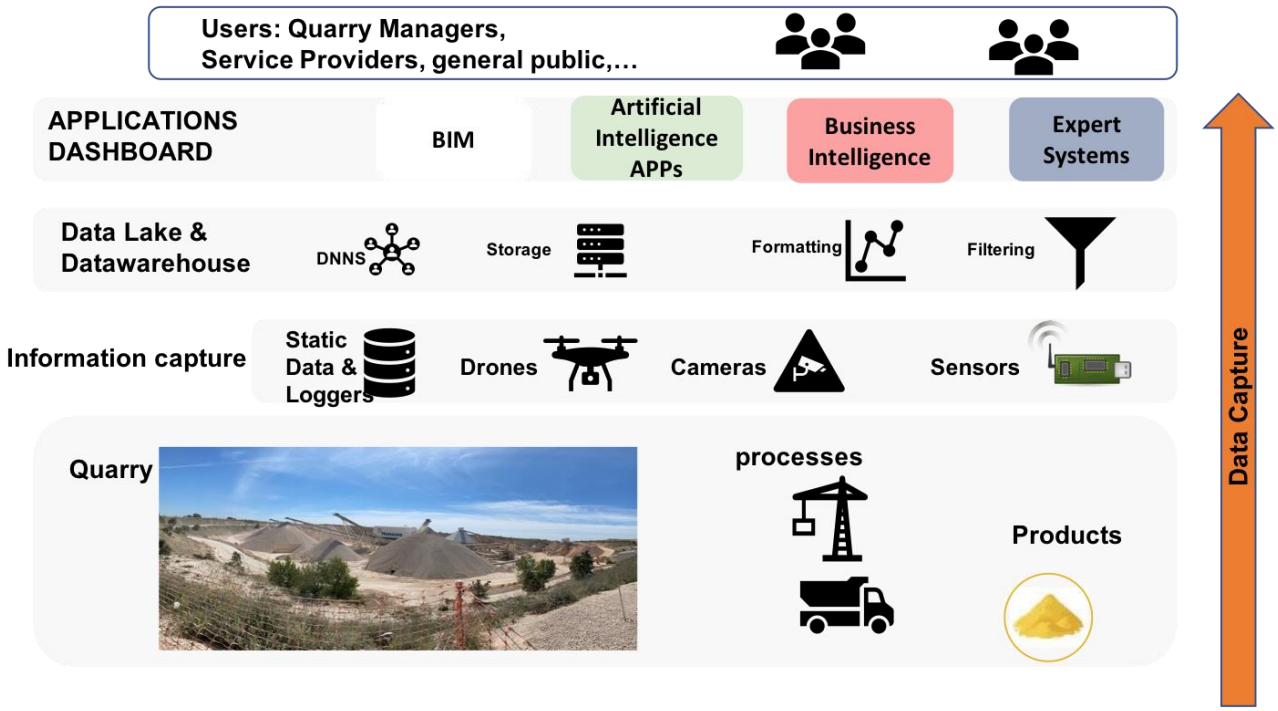

Figure 5: Data Architecture schema for DigiEcoQuarry

3.5.1.1 BIM

BIM (Building Information Modelling) is the process based on virtual model that simulates all aspects of the project and contains information on the characteristics of each building components. The BIM process itself last for all project lifecycle from design phase through the construction, management, maintenance, reconstruction, and demolishing. BIM is a very effective technology in terms of representing virtually a model that can be used for a very wide number of applications around almost any industrial activity.

In the case of DigiEcoQuarry, as of September 2021, the project is defining the functionalities that will be implemented within a BIM quarry model, in order to monitor the different processes of a quarry, as well as integrate disconnected systems and workflows that need to tie together in Common Data Environment (CDE). The ultimate goal is the production of Digital Twin which is a virtual representative of the project. The first step in terms of data processing is the extraction of the physical dimensions in 3D of the Quarry in order to build a model that can be used to access the different physical locations of a quarry (navigating a map), that could hold each a model that will deliver information about the processes data. Therefore, the BIM is a very handy user tool incorporating attractive and realistic quarry features in a form of a Digital Twin.

BIM maturity levels range from Level 0-3 and beyond, that defines what criteria are required to be consider in a specified way. The UK maturity model displayed in Figure 6: BIM maturity level. In the case of DigiEcoQuarry, BIM Level 3 will be implemented. BIM level 3 is a single collaborative, interoperable, online, project model with construction sequencing (4D), cost (5D), and project lifecycle information (6D and 7D) and is intended to deliver better business outcomes.

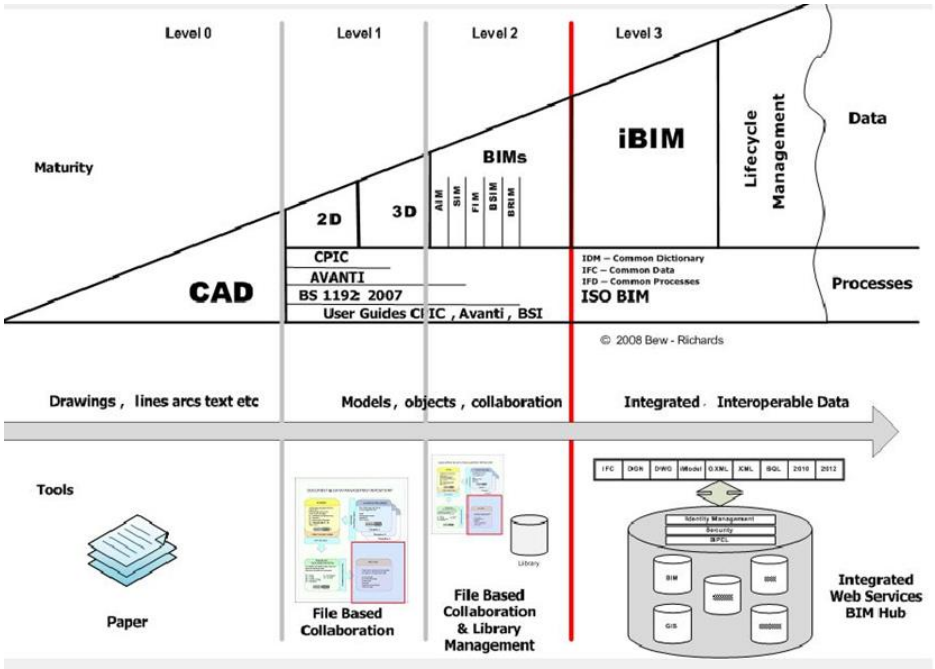

Figure 6: BIM Maturity level

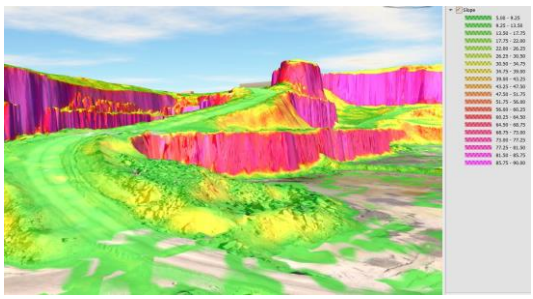

Figure 7: Example of a BIM model for a Quarry exploitation front

BIM is thus mainly in this case a data consumer that need information coming from different sources as displayed in Figure 4. As an example of the functionalities that could be introduced in the DEQ BIM, an extract of actual applications in the Building industry (https://www.buildingsmart.es) is presented below that focuses on the process management of a BIM model. The explanations were tuned to meet the quarry processes.

Table 2. BIM Features

Nr.	Service/Use Case	Explanation on the Data that is handled
1	**Record Modelling (RM)**	This has potential to contain information regarding not only the main architectural and physical elements, but equipment and asset information as well. The record model may contain information regarding design specifications allowing for validation that the as-constructed BIMs meets or exceeds these specifications. Furthermore, with the continuous improvement of the record model and the capability to store

Nr.	Service/Use Case	Explanation on the Data that is handled
		more information, the model contains a true representation of space with information such as serial numbers, warranties the components in the quarry.
2	As Constructed Modelling	A process in which BIMs contain an accurate depiction of the physical conditions and environment of a facility and its assets.
3	Asset Management (AM)	The assets included in BIM(s), consisting of the physical building, systems, surrounding environment, and equipment, must be maintained, upgraded, and operated at an efficiency which will satisfy both the owner and users at the lowest appropriate cost. It assists in financial decision-making, as well as short-term and long-term planning are also supported by the BIM(s).
4	Space Management and Tracking	A process in which BIM(s) are utilized to effectively allocate, manage, and track assigned workspaces and resources.
5	Maintenance & Repair Information	A process to allow the collection and storage of maintenance and repair information about objects to be brought into BIM(s).
6	Facility Management Documentation	A process where a facility data schema is developed to ensure information is supplied to the BIM(s) in electronic form so that it can be easily exchanged between the BIM(s) and the organizations selected IWMS tool.
7	Disaster Planning / Emergency Preparedness (DRC)	A process in which emergency responders would have access to critical building information in the form of BIM(s).The BIM(s) would provide critical building information to the responders, that would improve the efficiency of the response and, more importantly, minimize the safety risks.
8	Security Management	A process to identify and evaluate the security zones in the facility using BIM(s). The evaluation ensures that there are no "leaks" in the security strategy.
9	Communication move/add/change management (CM)	A process tying the BIM(s) to personnel management to help ensure that communications will be in place for all workers. This item also ensures that hard wired computers are operational when a person is in office space a quarry...
10	Way finding (WF)	A process of using the BIM(s) for documenting and aiding facility or site workers. Wayfinding encompasses all the ways in which people and machines orient themselves in the physical space and navigate from place to place.
11	Risk Management	A process to identify some issues that can be prevented by establishing data-driven analysis and simulations of potential disruptions. This analysis requires structured data sets provided by intelligent BIM(s)
12	Energy Efficiency	A process in which BIM can be used to help reduce annual energy use and minimize environmental risks
13	Common Data Environment (CDE)	A process of tying together various disconnected systems and workflows, and enables handover between all the participants in facility's lifecycle: architects, engineers, contractors, operators, facility managers, commissioning agents, etc.
14	Digital Twin	Refers to a digital replica of physical assets, processes, people, places, systems and devices that can be used for various purposes. The digital representation provides both the elements and the dynamics of how an Internet of Things device operates and lives throughout its life cycle.

Data will be fed to the BIM SW (IFC) normally through the data lake that will gather basically all data coming from the Quarries. Data within the BIM will need to be protected from unauthorized users and will need to follow strict data access and management standards as used for any web based application nowadays.

3.5.1.2 Expert Systems

The expert systems are digital systems able to generate, read and process data according to the area of expertise [e.g., at the processing plant, BIM, etc] intended to be analysed during the project. The idea is to use the systems and tools already developed by each partner, in case that these tools already exist, to avoid redundancies in analytics tools and to focus in the areas and/or results that needs to be developed. The expert systems will provide the necessary processed information for the creation of added value application between the different partners.

- Mobile Equipment Digitalization (2 different data models)

Table 3. Mobile Equipment Expert System Data Features (ABAUT)

PROVIDER		ABAUT SOLUTION
Nr.	Service/Use Case	Explanation on the Data that is handled
1	**Mobile Machinery –Loading & Transport (ABAUT)**	The traceability of the material will provide a deep knowledge of the quality and quantity of the material extracted for a correct processing and analysing at the processing plant storage, and internal transportation (the information can be crosscheck with the geological rock quality and material behaviour and grade, together with their impact on the other subsequent mining processes). To follow the mass-flow through out different sites in the quarry, it's necessary to refer to underline map giving information on the mining area (processing plant, extraction area, blasting fields, workshop, storage facilities, gas-station...) that would consequently correlate with the BIM of the mine. For the external transportation, is possible to follow the transport from the quarry to the customer. Data for customer, material, quantity may be retrieved from the central system or generated by visual logistic recognition services.
2	**Integrated management of the trucks fleet, including environment & safety aspects.**	The integrated fleet management system and analytics includes not only the availability of the machines, it is also connected with the machine utilization and their maintenance plant as well as with the production and daily plan. For increasing the performance of the fleet, it is possible to set different dashboards and analytics for the energy needed to carry out the different machine activities. This tool will improve the performance, reducing the energy needed and the environmental impact of the activities. Also, a correct maintenance plan and the implementation of best practices will reduce the supply chain regularity, the use of spare parts, and the energy and time needed to perform different machine activities can be reduced. In specific areas, stationary cameras can be installed for detecting the presence of people within the viewing range of the camera. Same applies for the cameras installed on the mobile mining equipment. To detect the presence of people and their proximity to machines, based on sensors.

		This report generation can bring a deep knowledge of the current activities and processes for a better decision-making step and operational performance. The integrated fleet management systems and analytics includes not only the availability of the machines, it is also connected with the machine utilization and their maintenance plan as well as with the production and daily plane.
3	**Automatic generation of the documents of the materials**	The automatic generation of documents and reports suppose an automation of certain activities carried out every day by the different operators (as an example working hours, production, activities report will save time for the operators and will also create standardization of the data for the different implementations in several other systems/services. This report generation can bring a deep knowledge of the current activities and processes for a better decision-making step and operational performance.

Table 4. Mobile Equipment Expert System Data Features (DH&P)

PROVIDER		DH&P
Nr.	Service/Use Case	Explanation on the Data that is handled
1	**Input data from mobile equipment (with focus on loading/hauling)**	The expert system smartQuarry relies on input from mobile machinery data to obtain multiple integrated analysis about the dynamic mining process and performance. The input data is gathered from sensors equipped directly by the equipment manufactures and as well as a set of additional retrofit sensors. The transmission and connection between sensors and a central installed data logger is managed with a CAN bus. The data is transmitted from the central data logger via mobile network or WLAN to a central server and IOT analyse platform. All data is provided with a time stamp that allows the full historical monitoring of the performance of each equipment. The expert system commonly records OEM pressure, flow and multiple engine and operator monitoring sensor values combined with retrofit DGPS, load cells, fuel sensors or tire pressure sensors etc. From there the basic machine outputs are monitored and displayed such as: ▪ Operating status ▪ Position ▪ Operating, working and idle times ▪ Utilization ▪ Driving distances, speeds, elevation travelled ▪ Fuel consumption ▪ Production performance (tonnage moved) ▪ Maintenance parameters (tire pressure, oil temperatures etc.)
2	**Integrated data of the fleet management, performance and material tracking**	All data gathered from the mobile equipment is processed and stored in a structured manner and integrated in the platform smartQuarry/AutoPLAN. The platform is a complex hub for data visualization, analysis and the data integration of additional inputs from the quarry management such as topographical and geological information. Focus must be laid on the combination of automatically obtained equipment data, manual inputs and existing models for the terrain and deposit.

		Combined with additional model data such as material types, 3D geofences and online material analysis data a detail material tracking from the mining face to the pre-crusher can be realized. In Combination with data from the treatment process the geological knowledge of the deposit is enhanced allowing a sustainable use of the specific resource and a controlled quality management. The integration of manual inputs together with the automatically obtained machine data allows also the adaptability of the system for example for maintenance management of the fleet. In the DEQ project the focus is on the control of the full production cycle, controlling and enhancing process performance, environmental and H&S benefits.
3	**Automatic generation of the documents of the materials**	The system offers multiple web services with personalized log-in. The services combine map visualization, customizable dash boards for machine parameter analysis and KPIs or fleet maintenance plans. The service includes a reporting service where customized reports can be downloaded or transmitted to defined parties on a regular basis. The system is customizable depending on the different needs of the specific quarry operation.

- **Treatment Plant Digitalization expert system**

Table 5. Treatment Plant Expert System Data Features (ARCO)

Nr.	Service/Use Case	Explanation on the Data that is handled
1	**Input data from machinery sensor.**	The expert system get data from the main machinery through the sensor installed. The sensor can work under analogical or digital system besides connect through Modbus. The most used types sensors are: Kilowatt meters, toroidal Transformers, load cell, radar and video including Frequency converters. This information will be: ▪ Energy measure: From individual important machinery as mill, screw, feeder, or main part of the treatment: primary, secondary, … ▪ Production: It is possible to know the commodity that each part produces through the weigh bridges, continuous weighs, radar or video. ▪ Waste: It is the same case that production. ▪ Time off, time load…
2	**Integrated data**	The different input data give a lot of information about the treatment process, production, performance, storage, energy, difference process time, safety even forecasts. Regarding this information, the expert system offers several outputs, for example: ▪ Production of different elements, consumed and important times. ▪ Control at all safety elements, warning, alerts, power drops, time down, maintenance, full/empty warehouse.

		It must be pointed the quantity of data that is necessary to save into the expert system and how much is stored in the data lake or IQS.
3	**Automatic generation of the documents of the materials.**	The system generate automatically Datasheet with statistics, it can work out data list too, which including time starting, time stopping, time down, power drops, performance, wastes, production… This information can be used by: ▪ Technical staff. ▪ Plant director. ▪ Managers. ▪ Suppler. ▪ Internal and external transport. ▪ Storage.

- **BIM Expert Systems**

Table 6. BIM Expert Systems Data Features (APP)

Nr.	Service/Use Case	Explanation on the Data that is handled
1	**Ecodomus – Project Lifecycle BIM**	EcoDomus is the leading expect system for project lifecycle BIM, that provides a 'Common Data Environment (CDE)' for all relevant types of information. It gives a 3D view of facilities in an easy-to-use format that links the Asset Information Model (in BIM) with real-time facility operations data acquired by meters and sensors (Building Management Systems, BMS) and facility management (FM) software. This allows for intelligent analysis of a building's performance and supports better maintenance practices, resulting in significant reductions in labour hours and energy use.
2	**Elecosoft - Powerproject BIM**	Powerproject is an Enterprise Project, Portfolio, Resource and Cost Management & 4D Planning software solution for organizations that manage people working across multiple projects, programmes and project portfolios.
3	**Trimble TILOS**	Linear project management solution provides a simplified, visual look at your construction project through a powerful linear scheduling view.
4	**Deltek - Acumen Suite**	Acumen Suite is a comprehensive Project Analytics, Diagnostics, Risk Analysis, Forecast Solutions and visualization and problem resolution platform.

- Safety Expert System for Processing Plants

Table 7. Safety Expert System Data Features (ITK)

Nr.	Service/Use Case	Explanation on the Data that is handled
1	**Safety Tool**	The safety management system for stationary equipment is a system that aims to reduce the risks of a collision between persons and machines and hence helps to achieve vision zero (zero accidents at construction sites) Based on the example of a scrapper system, sensors will be selected which shell detect objects (including machines and persons) coming close or interfering with the machine. The system shall then, based on rules (e.g. on classification rules) start measures to avoid harm (e.g. start with a noise/light alarm up until an emergency stop). Data to predict traffic (e.g. trucks) might come from a different expert system (e.g. 1 or 2).

3.5.1.3 Data lake, IoT Platform and Data warehouse

The Data lake will be a centralized and secured repository, by quarry, based on the cloud, to allow, in particular, the ingestion, the storage, the analysis and the usage of relevant data, available in its native format, such as structured or unstructured files but also databases, coming from each quarry (pilot sites) as well as from the technical providers partners working with them. These cloud Data lakes will be fully administrable to enable right accesses and give appropriate permissions. Data lake users (pilot site users, BIM, AI & Data warehouse, IoT devices, expert systems), once authenticated and authorized, will be able to:

- Create, Read, Update, Delete data (download/upload), according to Data Lake specifications
- Explore data lake content
- Share BIM, AI models and Expert system data
- Store:
  - Files, Images (txt, xml, csv, json, zip, jpeg, xls…) with associated metadata
  - Data Bases (SQL, NO SQL)
  - Events (IoT data, sensors data)

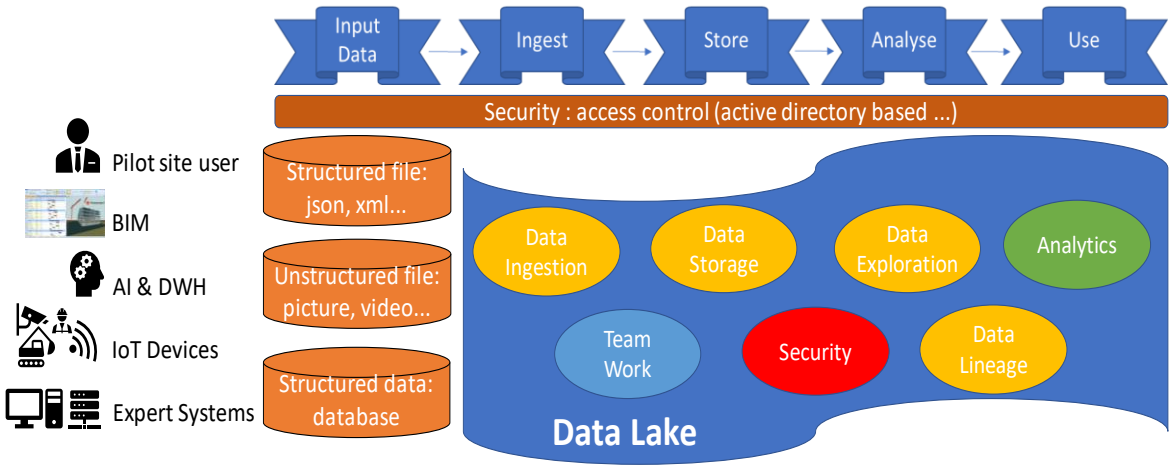

Figure 8: Data Lake global architecture

The IoT platform will build an IoT eco-system of objects of the physical world. It allows the digital representation of the devices on the quarry field. For that, it will gather data from the sensors and the expert systems. The IoT platform also will enable the development of applications using data from IoT data sources. These applications will cover all the mining processes: Blasting, Drilling, Loading & Transport, Storage, Treatment and Transport. The created applications will enable services that support the smart, sustainable and digital quarry. E.g. H&S, Efficiency, selectivity & profitability, Environmental Impact, Social Acceptance, Autonomous/remote operations, Process control, Advanced analytics, etc.

IoT data will be sent and shared securely in real time (if needed) and with low latency to generate insights and actions to help improve business and processes. The following picture illustrates the IoT core subsystems generally needed for the implementation of Thing, Insight and Actions concepts.

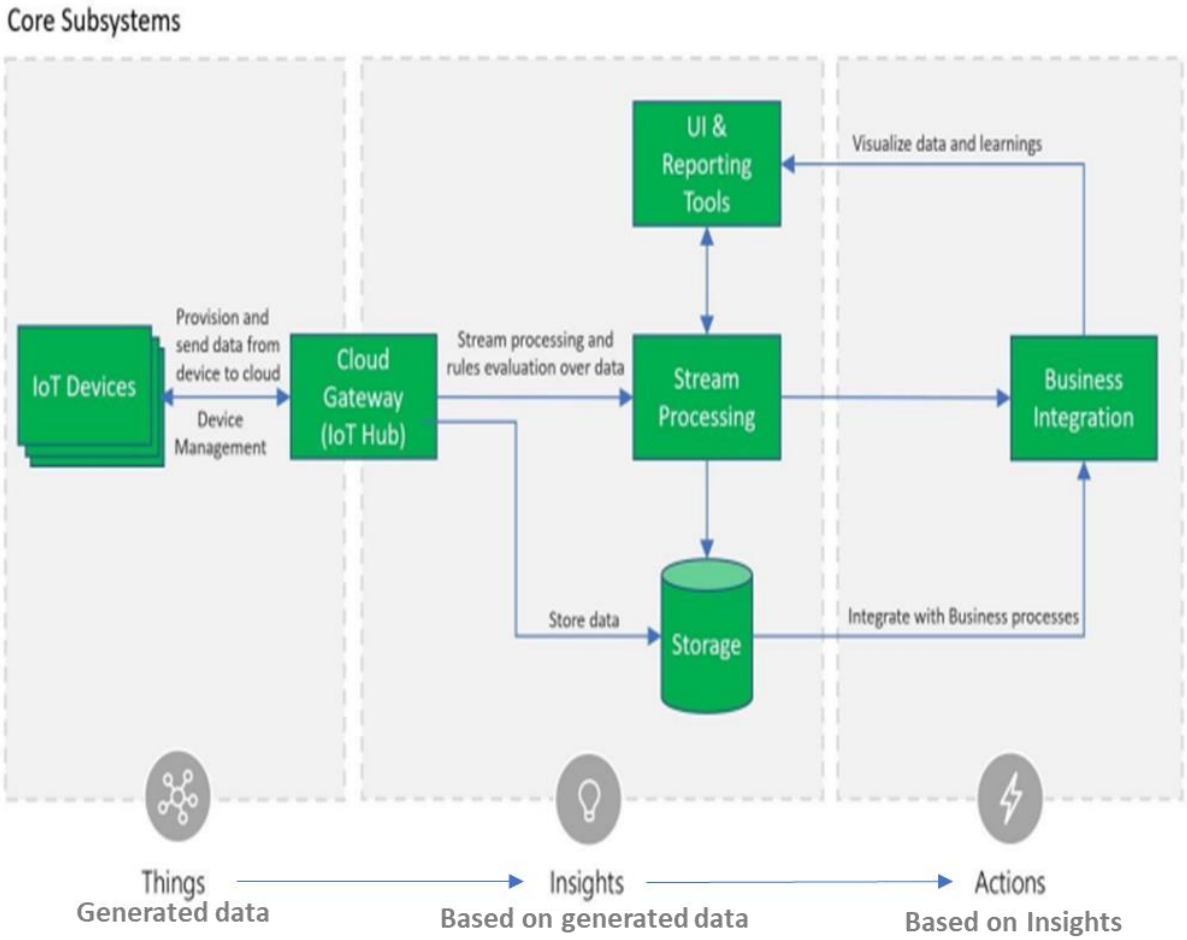

Figure 9: IoT Core subsystems and main principles

The Data warehouse mainly aims to analyze data. The Data warehouse will extract data from the Data lake that will be needed fundamentally for Business Intelligence and Artificial Intelligence services. The Data warehouse also will gather the results of data processing implemented by the AI algorithms: Data in this case will have a well-defined structure and organization in order to be retrieved by the applications dashboard easily following specific data construction models and APIs.

3.5.1.4 Artificial Intelligence algorithm applications

The Artificial Intelligence services that will be developed in the project will be fed with the data coming principally from the Data lake and Data warehouse framework as shown in Figure 5: Data Architecture schema for DigiEcoQuarry. The results of the algorithms will be also stored in the Data warehouse in such a way that the Digiecoquarry Application Dashboard composed of the IQS high level services.

The table below describes the AI services and explains what type of data and results are being pursued.

Table 8. Data characteristics tied to the services that will be developed using Artificial Intelligence algorithms

Nr.	Service/Use Case	Explanation on the Data that is handled
1	**Risks Detection (RIQ)**	▪ Risk identification and detection of a possible accident in the quarry characterizing the risk levels, possible locations. ▪ Inputs used to detect risks: Human factors, Weather, Environment and Equipment
2	**Metaquarry: Advanced information search on Quarries**	▪ Using PLN processing, the service via keywords will be able to look for specialized information on the quarry infrastructure, processes, business and workers. ▪ Data will be looked upon the datalake, specific quarry and equipment providers databases and the Internet in general. ▪ The data formats may comprise textual or any audio-visual type.
3	**PrMa: Predictive Maintenance (AM)**	▪ Using data coming from the quarry equipment sensors through the datalake, PrMA will be able to detect anomalies on the day to day operation of the quarry machines and equipment. ▪ PrMa will predict several equipment/machine characteristics such as future repair needs, end-of-life time intervals, possible solutions to a detected problem, etc.
4	**StockForecast: Stock handling and management functions**	The Stockforecast service will be handling production and market information in order to advice and inspect how the stock management of a quarry is being and should be performed.
5	**Hawkeye: Quarry visual inspection using AI image processing**	Hawkeye will basically image information provided by cameras, images coming from drones and satellite imaging in order to display quarry information about its surface analysing changes, providing guidance about material quality, point of extraction historical data and advices o future choices for points of extraction.

3.5.1.5 Business management tools

These tools require a very significant amount of information coming from all the infrastructure of Quarry and players. Specially for example the static information coming from each and every quarry providing detailed information about their business figures. As main interface the Business management tool will be using the Data warehouse to extract all needed information.

Table 9. Data characteristics tied to the services provided by the business management tool

Nr.	Service/Use Case	Explanation on the Data that is handled
1	**Aggregated data visualization**	Main quarry KPIs will be here accessed divided by the quarry processes: form the Site preparation activities till the Rehabilitation
2	**Energy Consumption**	Here all the quarry elements that are used for the production are considered and that may incur in energy consumption. The latter used for production specially.
3	**Load and Transport cycles**	Analysis of the product volumes handled per transport cycle
4	**Total inventory by site and forecast**	Here the number of different products handled in a quarry, their production levels forecasts and sales will be displayed
5	**Report by photography (mobile phone)**	Photos can be uploaded to the Datawarehouse form different points of the quarry by the workers in order to analyse possible risks, needed activities to be performed on the roads, maintenance of machines, etc…
6	**Productivity Assessment**	Productivity will be measured by showing a machine, factory or person efficiency in converting inputs into useful outputs.
7	**Work Environment & Worker Management**	Workforce detailed information will be here displayed in order to access easily statistic ABAUT KPIs that involve the workers performance, quarry working conditions, Health and security information, etc.
8	**Financial Management**	All important financial statistics are here accessed in terms of monetary transactions fundamentally.
9	**Quality Management**	Quality KPIs will be here displayed considering each process quality measures and how the quality controls are being performed in the quarry.
10	**Environmental Management**	A dashboard will be provided and will show basically the emission of all different gases and particles that are emitted within a quarry. Also, it will cover KPIs on how the environment is being taking care of within the quarry, displaying relevant statistics about waste, resources expenditure, handling of dangerous products, etc.
11	**Water Management**	Basically, the water consumption, use and recycling will be here analysed.
12	**Waste Management**	KPIs with respect of all sources of waste and how it is handled in a quarry is here displayed
13	**Maintenance Operations**	Here statistics about the Maintenance operation performed and needed can be observed, including costs.

4 DIGIECOQUARRY shared dataset description

4.1 Potential datasets / description

This section provides guidelines for the description of the different types of datasets to be collected and shared by DIGIECOQUARRY after the end of the project with respect to Open access to scientific publications which aims to improve and maximize access to and re-use of research data generated by Horizon 2020 projects and needed to validate the results presented in the deposited scientific publications . As the nature and extent of these datasets can evolve during the project, more detailed descriptions will be provided in the next version of the DMP (M47).

The descriptions of the different datasets, including their reference, file format, standards and metadata and archiving and preservation to be used are given below.

Table 10. DIGIECOQUARRY Dataset description template

Dataset Reference	Each dataset will have an identifier that will be generated by the combination of the name of the project, the pilot site, the use case in which it is generated and the datatype: "DIGIECOQUARRY_Pilot-Site_UC_Datatype".
Dataset Name	Name of the dataset.
Dataset Description	Each dataset will have a full data description explaining the data provenance, origin and usefulness. Reference may be made to existing data that could be reused.
Standards and metadata	The metadata attributes list to be used to explore the dataset.
File format / data format	Standard file type of the log file for example csv, xml, Jpeg Type of formatting of the payload of a log item, for example: binary, txt, json Structure and definition of log parameters that can be logged in a log item
Data Sharing (access right and License)	Explanation of the sharing policies related to the dataset between the next options: Open: Open for public disposal. Embargo: It will become public when the embargo period applied by the publisher is over. In case it is categorized as embargo, the end date of the embargo period must be written in DD/MM/YYYY format. Restricted: Only for project internal use. Each dataset must have its distribution license. Provide information about personal data and mention if the data is anonymized or not. Indicate if the dataset entails personal data and how this issue is taken into account. License: Example of type of licenses (cf. https://opendefinition.org/licenses/)
Archiving and Preservation	The preservation guarantee and the data storage during and after the project (for example databases, institutional repositories, public repositories, etc.).

5 Open access to scientific publications and underlying research data

DIGIECOQUARRY has agreed to provide open access to scientific publications and the underlying research data needed to validate the results presented in the deposited scientific publications according to article 29.2. DIGIECOQUARRY will use the specific Horizon 2020 guidelines associated with open access[1] to ensure that the results of the project provide the greatest impact possible.

DIGIECOQUARRY will ensure the open access to all peer-reviewed scientific publications relating to its results and will provide access to the research data needed to validate the results presented in deposited scientific publications.

The following lists the minimum fields of metadata that should come with a DIGIECOQUARRY project-generated scientific publication in a repository:

- The terms: “European Union (EU)”, “Horizon 2020”
- Name of the action (Research and Innovation Action)
- Acronym and grant number (DIGIECOQUARRY, 101003750)
- Publication date
- Length of embargo period if applicable
- Persistent identifier

When referencing open access data, DIGIECOQUARRY will include, at a minimum, the following statement demonstrating EU support (with relevant information included into the repository metadata):

“This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 101003750”.

In regard to the specific repositories available to the DIGIECOQUARRY consortium, some project partners maintain institutional repositories, where project scientific publications and in some instances, research data will be deposited. The use of a specific repository will depend primarily on the primary creator of the publication and on the data in question.

Other project partners will not operate publicly accessible institutional repositories. When depositing scientific publications, they shall use either a domain specific repository or use the EU recommended service OpenAIRE (http://www.openaire.eu/) as an initial step to finding resources to determine relevant repositories.

Project research data shall be deposited to the online data repository ZENODO . It is a free service developed by CERN under the EU FP7 project OpenAIRE plus (grant agreement no. 283595).

The repository shall also include information regarding the software, tools and instruments that were used by the dataset creator(s) so that secondary data users can access and then validate the results.

The DIGIECOQUARRY data collections will be accessed in ZENODO repository at a link similar to : https://zenodo.org/communities/<<DIGIECOQUARRY_REPOSITORY_NAME> to be defined.

1 https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

Besides, when partners of the DIGIECOQUARRY consortium are to participate in external events they must follow the guidelines described in Deliverable 9.1, sections 5.4.1 and 5.4.2. Each partner that publish a scientific publication has to inform ANEFA as leader of WP9 and follow section 5.3.4 guidelines of D9.1.

In summary, as a baseline, DIGIECOQUARRY beneficiaries shall deposit:

- Scientific publications – on their respective institute repositories in addition (when relevant) to the DIGIECOQUARRY ZENODO repository
- Research data needed to validate the results presented in the deposited scientific publications ('underlying data') to the DIGIECOQUARRY ZENODO repository.

Recently, the European Commission launched an open publication platform[2], for publishing scientific articles that will be accessible to all. The platform will showcase the results of research funded by Horizon Europe, the EU's research and innovation program for the period 2021-2027, and its predecessor, Horizon 2020. Open Research Europe will provide free access to the latest scientific discoveries for all, researchers and citizens alike. It provides a direct solution to the major difficulties often associated with publishing scientific results, including delays, barriers to re-use of results and high costs. The platform is an optional service offered to Horizon Europe and Horizon 2020 beneficiaries, so that they can meet their funding needs for immediate open access, at no cost to them. DIGIECOQUARRY beneficiaries will consider the usage of this platform or any other solution that will ensure the open access to all peer-reviewed scientific publications.

2 https://open-research-europe.ec.europa.eu/about

6 FAIR data management principles

In the context of Open-Access to scientific information described in article 29.2, the data underlying publications that will be shared should be FAIR [3], that is, Findable, Accessible, Interoperable and Reusable. These requirements do not affect implementation choices and do not necessarily suggest any specific technology, standard, or implementation solution.

The FAIR principles were generated to improve the practices for data management and data curation, and FAIR aims to describe the principles in order to be applied to a wide range of data management purposes, whether it is data collection or data management of larger research projects regardless of scientific disciplines.

With the endorsement of the FAIR principles by H2020 and their implementation in the guidelines for H2020, the FAIR principles serve as a template for data lifecycle management and ensure that the most important components for lifecycle are covered.

This is intended as an implementation of the FAIR concept rather than a strict technical implementation of the FAIR principles. DIGIECOQUARRY will implement several actions described below to carry on the FAIR principles.

Making data findable, including provisions for metadata:

- The datasets will have very rich metadata to facilitate the findability. Open data format (csv, xml) will be used.
- All the datasets will have Digital Object Identifiers provided by the public repository (ZENODO).
- The reference used for the dataset will follow a format like the following: "DIGIECOQUARRY_PilotSite_UC_Datatype_XX" (XX: identifier to be added for similar datasets).

Making data openly accessible:

- The datasets that will be openly available will be described according to Table 10.
- The datasets will be made available using a public repository (e.g. ZENODO) after the project.
- Table 10 will be used to explain the methods or software used to access the data. Basically, no software is needed to access the data.
- The data and their associated metadata will be deposed in a public repository or either in an institutional repository.
- The Data Sharing row Table 10 will outline the rules to access the data if restrictions exist.

Making data interoperable:

- Open and widely used data formats in quarry sector (csv, jpeg, IFC, xml, json) will be used. This list will be updated.

Increase data re-use (through clarifying licenses):

- The data producers will make their data available for third parties within public repositories. The data will be reused for the scientific publications' validation purpose.

3 https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

- Data producers can license their data to allow the widest reuse possible. Examples of licenses are available at https://opendefinition.org/licenses/ .

7 Allocation of resources

The costs to make the data FAIR in DIGIECOQUARRY shall be handled by each partner willing to generate and share its data.

In the project, each beneficiary should have a Data Manager who liaises with the PTC (Project Technical Coordination) /WPT (Work Package Team) about data management issues.

Scientific publications and the underlying research data collected as part of this project and needed to validate the results are owned by the data producers or partners involved in the pilot sites. The partners take the responsibility for the collection, management, and sharing of the research data. Data quality assessment will be the responsibility of each pilot site.

8 Data Protection Regulation and Ethical Aspects

8.1 GDPR Application in DIGIECOQUARRY

8.1.1 Background

Data protection is both a central issue for research ethics in Europe and a fundamental human right. It is intimately linked to autonomy and human dignity, and the principle that everyone should be valued and respected.

The protection of personal data in the European Union is guaranteed by Article 8 of the EU Charter of Fundamental Rights (CFR) and Article 16 of the Treaty on the Functioning of the European Union (TFEU).

In the era of technical development progress and innovations, more and more information and personal data is digitalized and can be retrieved more easily and quickly. Therefore, the informational self-determination of EU citizens is both important and challenging. The abundance of information and the ease of its storage and combination, awakes desires and carries high risks. Data protection is intended to counter these risks preventively.

To meet the new challenges, the EU recently has introduced the General Data Protection Regulation (GDPR) which is directly applicable in the EU countries since 2018 and replaces the previous Data Protection Directive.

DIGIECOQUARRY will develop systems, technologies and processes for integrated digitisation and automation real-time process control, to be piloted in 5 EU quarries with the target of improving health and safety conditions for workers. The pilot tests will lead to improved efficiency of processes, maximizing quarry resources and sustainable management of water, energy emissions, minimised environmental impact and expanding the EU aggregates and construction business. Coupling Artificial Intelligence approaches with cyber-physical systems and the Internet of Things concept, will make Industry 4.0 approach possible and the smart sustainable extractive site a reality.

To achieve the research and development of such integrated digitisation and automation real-time process control systems and processes, personal data will have to be collected and processed.

The DEQ Consortium is fully aware of this and its related challenges in terms of research ethics and, more specifically, data protection.

In the next section we will analyse the essential elements of the GDPR for its application in the DIGIECOQUARRY project (DEQ).

8.1.2 Key concepts and definitions

The GDPR defines some terms that are relevant to data protection in DEQ:

- **Personal data** means any information relating to an identified or identifiable natural person who can be identified, directly or indirectly, by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person (Art. 4 GDPR).
- **Special categories of personal data** are personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership [...] as well as genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation (Art. 9 GDPR).
- **Processing** means any operation or set of operations which is performed on personal data or on sets of personal data, whether by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure, or destruction (Art. 4 GDPR).
- **Profiling** means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location, or movements (Art. 4 GDPR).
- **Anonymisation:** a process of ensuring that the risk of somebody being identified in the data is negligible.
- **Pseudonymisation** means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person (Art. 4 GDPR).
- **Risk based approach:** data protection must be proportionate to the risks to data subjects.
- **Consent of the data subject** means any freely given, specific, informed, and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.
- **Controller** means the natural or legal person, public authority, agency, or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data [...] (Art. 4 GDPR).
- **Processor** means a natural or legal person, public authority, agency, or other body which processes personal data on behalf of the controller (Art. 4 GDPR).

8.1.3 Data protection principles

Regarding data protection according to GDPR, there are seven basic principles that must be respected when collecting and processing personal data in DEQ (see Art. 5 GDPR):

Figure 10: Basic principles of the GDPR

Personal data must be processed lawfully. The consent to the storage and processing of personal data must be voluntary, specific, unambiguous, and based on information. Personal data must be processed fairly and, in a manner understandable to the data subject. The principle of transparency assumes that all information is easily accessible and understandable and written in clear and simple language.

- Purpose Limitation

Personal data must be collected for specified, explicit and legitimate purposes and must not be further processed in a way incompatible with those purposes. Further processing for archiving scientific research purposes shall not be incompatible with the initial purposes if it is in accordance with Article 89.1 GDPR.

- Data Minimisation

Personal data must be adequate, relevant, and limited to what is necessary for the purposes of the processing. Personal data should be allowed to be processed only if the purpose of the processing cannot reasonably be achieved by other means.

- Accuracy

Personal data must be accurate and, where necessary, kept up to date. All reasonable steps shall be taken to ensure that personal data which are inaccurate for the purposes of their processing are erased or rectified without delay.

- Storage Limitation

Personal data must be stored so that it allows data subjects not to be identified longer than necessary for the purposes for which they are processed. Exception: Personal data may be retained for longer periods, to the extent that the personal data are used exclusively for scientific and non-commercial purposes in accordance with Article 89.1 GDPR subject to the implementation of appropriate technical and organisational measures required by the GPDR to protect the rights and freedoms of the data subject.

- Integrity and Confidentiality

Personal data must be processed in a manner which ensures adequate security of personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage using appropriate technical or organisational measures.

- Accountability

The controller shall be responsible for and be able to prove compliance with the above principles and GDPR requirements.

8.2 An approach to comply with GDPR in DIGIECOQUARRY

8.2.1 Data Usage in DEQ

All data processed within the framework of DEQ comply with the data protection principles described above. All beneficiaries involved in DEQ will adhere to these principles. DEQ will essentially collect mandatory data to support its research activities.

All consortium members have agreed that DEQ participants' personal data will be safeguarded as follows:

- No unnecessary data collection activities will be performed within the project's lifetime.
- Personal data that have been collected for analysis will be handled discretely and with anonymity.
- No personal data will be collected without definite permission of the human participants.
- Every personal data collected throughout the project will be treated with respect to the protection of fundamental human rights (e.g., separating general and personal data, handling encrypted personal data and identities, erasing irrelevant personal data).
- The participants will be granted with the right to access their personal data and the analysis and user models made based on it.

Compliance with data protection principles such as data minimisation should be guaranteed prior to data collection and processing.

8.2.2 Data Protection in Collection and Processing

Within the DEQ project, there are several partners who collect and process personal data of different kinds and from different sources. Each pilot site should name its own Data controller and data processor in order to collect and process the data.

Controllers are the main decision-makers. They exercise overall control over the purposes and means of the processing of personal data. Processors act on behalf of, and only on the instructions of the controller. The processor should only process personal data in line with the controller's instructions.

The controller must comply with and prove compliance with all the data protection principles. He or she will be responsible for the compliance of the processor.

It is important that each partner is individually responsible for compliance with current legislation. This includes their personal responsibility to carry out their tasks in a way that ensures compliance with the DEQ approach of data protection by design and by default.

In some cases, it is likely that the processing of personal data in DEQ raises risks concerning ethics and data protection, e.g., the processing of sensitive health data of employees or the usage of new technologies as automated data collection with electronic devices, profiling, tracking or the combination of data from different sources.

To counter and mitigate these risks, ethical and legal issues must be identified and analyzed. A questionnaire on ethics and data protection will be sent to each partner so that they can identify these aspects. The template for this questionnaire will be available as an annex to D 11.1.

Appropriate technical and organizational measures must be taken in advance of data collection and processing. These could be for example:

- the pseudonymisation or anonymisation of personal data.
- data minimisation.
- applied cryptography (e.g., encryption and hashing).
- using data-protection focused service providers and storage platforms. Arrangements that enable data subjects to exercise their fundamental rights, e.g., access to their personal data and consent to its use or transfer.

Each data controller must keep a record of the processing activities for which he or she is responsible (accountability principle) including adequate documentation on what personal data are processed, how, to what purpose, how long; documented processes and procedures aiming at tackling data protection issues at an early state when building information systems or responding to a data breach; the purposes of the processing, a description of the categories of data and recipients of the data and a general description of the technical and organizational safeguard measures. The documentation obligations depend on the risks posed by the nature of the processing of personal data.

In accordance with the provisions of article 30 of the GDPR, the basic content of the records to be considered compliant with the documentation and accountability obligations, must have at least the following information:

- Name and contact details of the controller and, where applicable, the joint controller, the controller's representative and the data protection officer;
- Purposes of the processing;
- Description of the categories of data subjects and of the categories of personal data;

- The categories of recipients to whom the personal data have been or will be disclosed including recipients in third countries or international organisations;
- Where applicable, transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation and, in the case of transfers referred to in the second subparagraph of Article 49(1) GDPR , the documentation of suitable safeguards;
- Where possible, the envisaged time limits for erasure of the different categories of data;
- here possible, a general description of the technical and organisational security measures referred to in Article 32(1) GDPR.

The controllers and processors are provided with some record templates in **Annex 1.**

Different methods of assessing the impact on data protection and privacy could be used to support the implementation of the essential requirements laid down in the GDPR.

To assure data protection and security of the research and to reduce risks and discomfort for the participants, necessary measures must be implemented which contain, inter alia, the following points:

- **Informed consent:** Informed consent is required for data collection, data storage, data processing and publication of raw or processed data. Before consent is sought, information must be given, specifying the alternatives, risks, and benefits for those involved, in a way users understand.
- **Voluntary participation:** Participation is on a voluntary basis.
- **Minimal risk:** Participants should not be exposed to more than minimal risk. The data would be retained for the shortest period necessary to achieve the project objectives.
- **Anonymity:** Volunteers have the right to remain anonymous. All data analyses are performed on an anonymous basis. It is very difficult to create completely anonymous datasets that contain the detailed information needed for research purposes. It also may be important to have the possibility to re-identify the research subjects to their personal data. In these cases, data should be pseudonymised to protect the data subject's privacy and minimise the risk to their fundamental rights in the event of unauthorized access. Therefore, all participants should be assigned an identification code so that the different actions of the participants during data collection can be mapped. The relationship between the identification code and the participant is captured in the repository and stored separately and securely. This file is accessible only to the data controller.
- **Feedback:** Participants shall be provided with the possibility to retrieve feedback on the results of research.
- **Privacy:** Researchers must ensure that the way research outcomes are reported does not contravene the right to privacy and data protection.
- **Confidentiality:** It is different from the participant's right to privacy; it refers to how data about the participants will be stored.
- **Data control:** The data subject has the right to access all data processed about him or her, and has the right to demand the rectification, deletion or blocking of data that is incomplete, inaccurate or isn't being processed in compliance with the data protection rules.
- **Informed stakeholders:** Informing stakeholders in detail on ethical aspects of research and evaluation/validation in reporting activities.

The Data Controller must take responsibility for ensuring that training procedures, supervision, and data security arrangements are sufficient to prevent unauthorized breaches of confidentiality.

Each access to the collected data is password protected and is only granted to authorized partners for data processing.

In addition, access to information will also be restricted to the partners involved in the respective task to ensure their confidentiality. Information technologies will be used to securely store, provide, and access data and to manage the rights of users.

8.2.3 Data Protection during Storage and Destruction

- Data from electronic records would be stored on secure servers that would be backed up weekly.
- In addition, technical measures would be in place such as firewalls, regularly updated virus protection and access restriction by password protection.
- Non-electronic files would be stored in locked file cabinets in non-public areas.
- Files containing information that could identify a study participant would be kept separate from research files in locked cabinets.
- All areas would be reviewed regularly to ensure that confidential information is not omitted in open areas.
- The data subject has the right to ensure that his or her personal data are erased and no longer processed if he or she has withdrawn the consent to the processing.
- All data from the project pilots and studies considered confidential will be discarded by the end of the project, while only the public models and respective datasets described in detail in the Section 4.1 Potential datasets / description will be made public.
- Several critical factors relevant to data retention will be considered, e.g., the purpose of data retention, the type of open data collected, the policies for access to open data, security measures, confidentiality, and anonymity of data.
- Existing methods and tools for permanent and irreversible data destruction would be used, e.g., completely overwriting, and formatting hard disks.
- Further processing of the personal data collected previously ("secondary use") is not envisaged. In the unlikely event that such a need arises during the project, the consortium will take all necessary steps to ensure explicit confirmation that the beneficiary has a legitimate basis for processing and that appropriate technical and organisational measures have been taken to protect the rights of data subjects. In addition, it will be ensured that the secondary data used in DEQ remain anonymous.

8.2.4 Data Protection Risk Assessment

The assessment of the data security risk should consider the risks arising from the processing of personal data, such as accidental or unlawful destruction, loss, alteration, unauthorised disclosure, or access to personal data transmitted, stored, or otherwise processed which may lead to physical, material, or non-material damage.

Where the processing may present a high risk to the rights and freedoms of natural persons, the controller should be responsible for the implementation of a DPIA, which shall assess the origin, nature, particularity, and

severity of that risk. The outcome of the assessment should be considered when determining the appropriate technical and organisational measures.

If a DPIA finds that processing operations pose a high risk that the controller cannot mitigate by appropriate measures in relation to the available technology and costs of implementation, the supervisory authority should be consulted prior to the processing.

In general, a DPIA concerns a single data processing operation. A single assessment may address a set of similar processing operations that present similar high risks. The controller shall have recourse to the assistance of the DPO where necessary.

The assessment shall contain at least:

- a systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller.
- an assessment of the necessity and proportionality of the processing operations in relation to the purposes.
- an assessment of the risks to the rights and freedoms of data subjects.
- the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with the GDPR considering the rights and legitimate interests of data subjects and other persons concerned.

The controller shall seek the advice of the DPO, when carrying out a DPIA. The DPO may provide advice and monitor the performance of the DPIA. In addition, the controller shall seek the views of the data subjects or their representatives on the intended processing, without prejudice to the protection of commercial or public interests or the security of the processing, where appropriate.

The legislation does not specify the method by which a DPIA is to be conducted, but there are criteria that it must include. In the figure bellow, the general process to carry out a DPIA is explained:

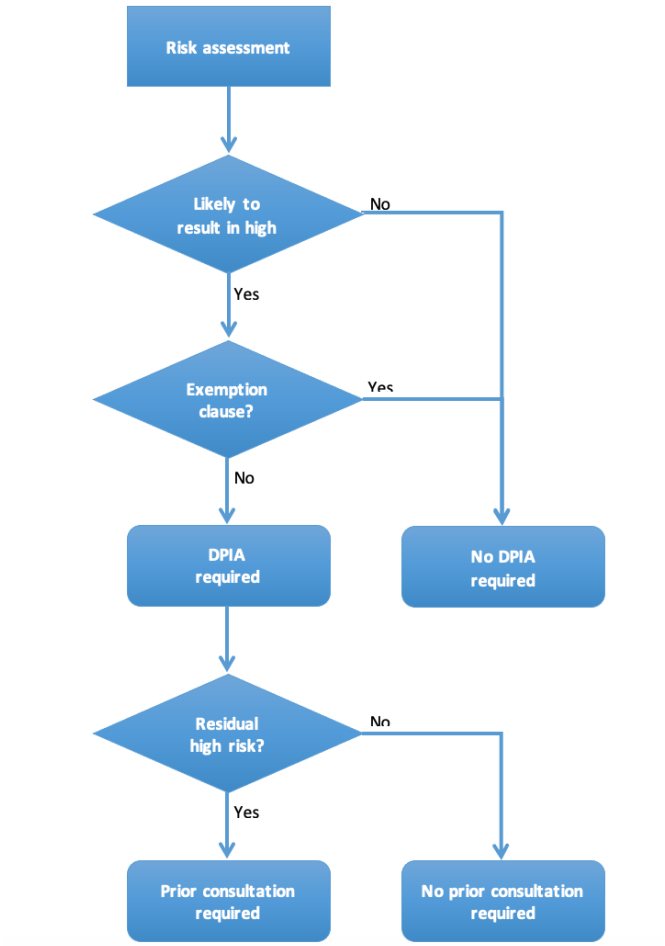

Figure 11 General process to carry out a DPIA

Partners will be provided with a DPIA template, set put in the European Guidelines.

8.3 Data protection regulation in the non-EU Countries

Although the majority of the DIGIECOQUARRY project will be developed within the European area by partners based in countries of the European Union to which the GDPR will be directly applicable, two entities not belonging to EU member states will participate in the project: The Colombian Association of Petro Aggregates Producers (ASOGRAVAS) and MinTek Industries South Africa (Pty) Ltd.

Regarding the regulatory framework for the protection of personal data in both countries, the main applicable regulations in Colombia and South Africa are:

8.3.1 Colombia and the General Data Protection Law

Since 2012, Colombian legislation has regulated Statutory Law 1581, which functions as a framework for the General Data Protection Law in the country.

This Law recognizes and protects the right that all people must know, update and rectify the information that has been collected about them in databases or files susceptible to treatment by public or private entities.

Thanks to this milestone, for example, companies have the obligation to request authorization on the destination and use of the data managed by their users.

In accordance with Colombian legislation, if an improper processing of personal data is identified, Law 1581 suggests resorting to the person directly responsible. However, in situations of total silence, it is possible to appeal to the Superintendency of Industry and Commerce (SIC), an entity that has the obligation to investigate and generate the corresponding fine in case of verifying any procedural anomaly.

8.3.2 South Africa and Protection Personal Information Act (POPI Act)-POPIA

The Personal Data Protection Act 4 of 2013 (POPIA), effective July 1, 2020, is South Africa's first comprehensive law dealing primarily with data protection. Its objective is to make the constitutional right to privacy effective by introducing measures that regulate the collection, treatment and use of personal information by public and private organizations in a fair, transparent, and secure manner.

POPIA's goal is to align South Africa with global data protection legislation and best practices, and in many cases is compliant with Europe's General Data Protection Regulation (GDPR). Since POPIA became law in 2013, there have been various advances, such as the appointment of the Information Regulator in 2016 and the publication of the definitive Regulation in 2018Without prejudice to the applicable regulations, these countries have not yet been declared by the Commission as countries with an adequate level of protection. That is why it is mandatory to adopt a series of measures in the processing of data within the DEQ project to ensure that the principles and rights described in the previous sections are scrupulously complied with by each of the project partners without diminishing the rights of the subjects.

When personal data is transferred outside the EU, the protection offered by the GDPR will accompany the data. That means that if the data is exported abroad, the controller must ensure that one of the following conditions is met:

- The subject has explicitly given his/her consent.
- The transfer is necessary for the execution of the project.
- The data protection of the non-EU member country is considered adequate.
- Necessary measures are taken to provide the appropriate safeguards, such as the inclusion of specific clauses in the contract concluded with the non-European importer of personal data.

When neither of these exceptions is applicable, the transfer can only be carried out if it is not repetitive, affects only a limited number of interested persons, is necessary for the purposes of compelling legitimate interests pursued by the person responsible for the treatment on which the interests or rights and freedoms of the person concerned do not prevail, and the controller evaluates all the concurrent circumstances in the data transfer and, based on this evaluation, offers appropriate guarantees regarding the protection of personal data.

In this case, the controller will inform the transfer control authority. In addition to the information referred to in articles 13 and 14 of the GDPR, the controller will inform the interested person of the transfer and of the compelling legitimate interests pursued.

In case personal data is transferred from the EU to a non-EU country, the controller must provide details on the type of personal data to be exported. The Controller must confirm that such transfers are in accordance with the Chapter V of the GDPR.

In case personal data are transferred form a non-EU country to the EU, the Controller must provide details on the type of personal data to be imported. The Controller must also confirm that such transfers comply with the laws of the countries in which the data was collected.

8.3.3 IPR

The IPR and knowledge management strategy are defined in D9.1. Dissemination, Communication and Exploitation Plan.

This strategy for the DIGIECOQUARRY will be designed according to the background of each partner, the ownership of the foreground identified and the exploitation agreements among the parties In line with this DMP.

Most of the research data will be confidential within the partners at least until the commercialisation starts or until the partners protect the results by new or already identified IPR tools. Some of the data as the raw data will not be shared by the partners that generates this information due to IP protection [e.g., patent protection, copy protection, etc.]. The information shared, already agreed in CA, will be processed data, the outputs of the different expert systems and of the different analytics software and systems

9 Ethical aspects

An Ethics / Privacy by Design approach is pursued within the framework of the project. This means that ethical and data protection issues are to be considered and observed during the planning and implementation phase for the entire duration of the project and beyond. This approach will be explained in further detail in the D11.1. "Ethics Requirements". In any case, the ethical strategy for the DIGIECOQUARRY project will be briefly explained at this point:

- All participants in the project are committed to ensuring their adherence to the ethical principles and to acting in compliance with the laws and codes of conduct.
- Particular attention must be paid to data protection, compliance with the requirements of the GDPR as well as protection of confidentiality and anonymity of participants involved.
- Each partner is responsible for acting in accordance with this rule. To ensure the same ethical and data protection guarantees, all DEQ partners must sign a document of compliance to the provisions contained in the GDPR, including those belonging to non-European states.
- An Informed Consent Form (ICF) will be prepared. The participants care to be informed about, among other things, the objectives of the project, the persons responsible, the methods used, the data collected, the handling and processing of this data, possible benefits, and risks as well as the possibility of withdrawing their participation.
- Partners involved in the project should name an Ethics Responsible Person (ERB). This person is responsible for ensuring that all activities of the respective WP take place in conformity with applicable law, that basic ethical principles are observed and that the anonymity of the participants and the confidentiality of their data are protected.
- An Ethics Advisory Board (EAB), formed by the Ethics Responsible Persons, will support the beneficiaries helping and advising the partners in ensuring compliance with ethical principles, the rights of study participants and conformity with applicable law in the areas of ethics, data protection and related issues.
- Risks to the rights and freedoms of natural persons may result from personal data processing which could lead to physical, material, or non-material damage. DEQ Partners will receive an ethical

questionnaire to classify the type of personal data that will be processed to evaluate risks based on an objective assessment. The severity and likelihood of the risks to the rights and freedoms of data subjects would be determined by reference to the nature, scope, context, and purposes of the data processing.

- If a high risk is detected, the controller, with the assistance of the DPO, must carry out a DPIA.

10 Data Security

The data produced during the execution of DIGIECOQUARRY project will be stored per each pilot site in dedicated servers depending of the use case and expert system for the whole project. According to the previous section, those data are made compliant with the GDPR. This section describes some security principles that are implemented in order to protect against any type of modification. Also, a more thorough management of the servers could be made using the ISO 27001 standards or the BSI-Standards. The security principles are listed below:

- Authentication: All the users wanting to get access to the DIGIECOQUARRY data servers should be authenticated. Also, proper means are used to authenticate the servers. An authentication system will be used to handle the authentication of the users during the course of the project.
- Authorization: The access to DIGIECOQUARRY data servers is only available to the authenticated and authorized users. These categories and the rights of those users are defined and enforced. The appropriate access control policies and mechanisms (including physical access control) shall be identified for each trial site and project wide to provide the authorization.
- Accounting: In DIGIECOQUARRY any access and modification to a resource by any user is securely logged in order to prevent users from denying that data files were accessed, altered or deleted, when auditing.
- Confidentiality: The data stored in DIGIECOQUARRY servers should be encrypted during transmission and storage.
- Communication Security: Access to DIGIECOQUARRY servers should be done through encrypted communication channels such as HTTPS and IPsec.
- Data Integrity: The data collected during DIGIECOQUARRY should be protected from malicious and accidental modifications by any users during their transmission or their storage. Cryptographic mechanisms such as hash functions and digital signatures shall be used.
- Availability: This security principle assures that the DIGIECOQUARRY servers should be available for DIGIECOQUARRY users during the defined interval of service. Also, regular backups of the data should be made.

11 Conclusion

This document describes the main principles and guidelines for the Data Management for the DIGIECOQUARRY project.

As living document, it is highly probable that as the project develops, eventualities will arise that make it necessary to update it. Changes to the datasets may have been made after more comprehensive studies of the pilot sites have been conducted. The next deliverable is set to month 47, "D. 10.11 Data Management Plan 2/2".

Next actions will be to focus on semantics and further clarification of procedures and identifying areas that need special attention.

The DIGIECOQUARRY DMP will put a strong emphasis of the appropriate collection and publication of data, storing all the information necessary for the optimal use and reuse of those datasets.

Once the solutions to be tested by the project and the content of this testing are specified, the next step – which will be described in the updated version of this report due in month 47 – will be to finalize the specifications of the DIGIECOQUARRY DMP.

12 Annexes

12.1 Annex 1: Data management organisation

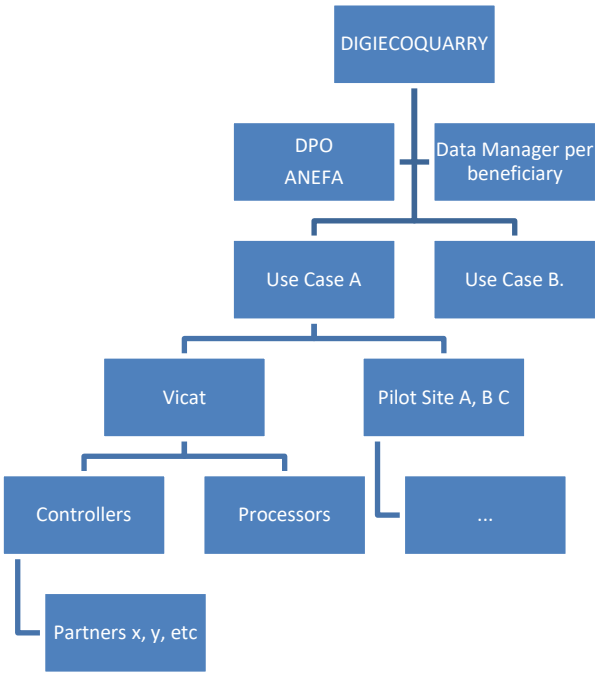

Figure 12: Data management organisation

12.2 Annex 2: Basic record template: Processing record example

12.2.1 Template of a processing record of the controller established in the EU

The following information are the legally required minimum requirements, which shall be made available to the supervisory authority on request (article 30.4 of the GDPR).

Table 11. Template of a processing record of the controller established in the EU

Name and address of the Controller				Data Protection Officer			
Name: E-mail: Tel.: Address:				Name: E-mail: Tel:			
nº	Purpose	Group Concerned	Category of data	Addressee	Transfer to third party	Erasure time	Measures

12.2.2 Template of a processing record of the processor established in the EU

The processor is also required to maintain processing records, but the information requirements differ.

Table 12. Template of a processing record of the processor established in the EU

Name and address of the processor:	Data Protection Officer:
Nº	
Controller	Name/address/e-mail/tel.
Category of processing	Abstract description of the service or services
Transfer to third countries	If applicable
Measures according to Art. 32 GDPR	Ex. Measures according to the safety concept+ additional encryption if agreed upon with the controller

References

Table 13. References

Document Resource ID	Document Resource name and reference
DR1	EU Grant Agreement n°101003750

List of Figures

List of Tables