MATRYCS

D2.2 | MATRYCS Technical and Security Specification

WP2 – System Requirements and Specifications

April 2021

www.matrycs.eu

Modular Big Data Applications for Holistic Energy Services in Buildings

Disclaimer

Copyright Message

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

Grant Agreement Number	101000158	Acronym	MATRYCS
Full Title	Modular Big Data Applications for Holistic Energy Services in Buildings		
Topic	LC-SC3-B4E-6-2020 | Big data for buildings		
Funding scheme	H2020- IA: Innovation Action		
Start Date	October 2020	Duration	36
Project URL	www.matrycs.eu		
Project Coordinator	ENG		
Deliverable	D2.2 MATRYCS Technical and Security Specification		
Work Package	WP2 – System Requirements and Specifications		
Delivery Month (DoA)	M7	Version	1.0
Actual Delivery Date	30/04/2021		
Nature	Report	Dissemination Level	Public
Lead Beneficiary	NTUA		
Authors	Miha Smolnikar [COMSENSUS], David Carro [COMSENSUS], Dario Pellegrino [ENG], Leandro Lombardo [ENG], Francesco Nucci [ENG], Pasquale Andriani [ENG], Marco Pau [RWTH], Zhiyu Pan [RWTH], Federico Carere [ASM], Tommaso Bragatto [ASM], Agnieszka Łukaszewska [FASADA], Marek Giluń [FASADA], Marina Bartolomé Lorenzo [VEOLIA], Jaime Gómez Tribiño [VEOLIA], Danielle Antonucci [EURAC], Sofía Mulero Palencia [CARTIF], Víctor Iván Serna [CARTIF], Javier Antolín [CARTIF], Roberto Sanz [CARTIF], José María Llamas [CARTIF], Carla Rodríguez Alonso [CARTIF], Gema Hernández Moral [CARTIF], Elissaios Sarmas [NTUA], Vaggelis Marinakis [NTUA], Haris Doukas [NTUA], Zoi Mylona [HOLISTIC]		
Quality Reviewer(s):	David Carro Santomé [COMSENSUS}, Alice Pittini [HOUSING EUROPE]		
Keywords	Big Data Value Chain, Technical Specification, User Requirements, Technical Requirements, Security Requirements, Data protection, IPR, Ethical requirements		

Preface

MATRYCS focuses on addressing emerging challenges in Big Data management for buildings with an **open holistic solution** for Business-to-Business platforms, able to give a competitive solution to stakeholders operating in building sector and to open new market opportunities. **MATRYCS Modular Toolbox,** will realise a holistic, state-of-the-art AI-empowered framework for decision-support models, data analytics and visualisations for Digital Building Twins and real-life applications aiming to have

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

significant impact on the building sector and its lifecycle, as it will have the ability to be utilised in a wide range of use cases under different perspectives:

- Monitoring and improvement of the energy performance of buildings - **MATRYCS-PERFORMANCE**
- Design facilitation and development of building infrastructure - **MATRYCS-DESIGN**
- Policy making support and policy impact assessment - **MATRYCS-POLICY**
- De-risking of investments in energy efficiency - **MATRYCS-FUND**

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

Who We Are

	Participant Name	Short Name	Country Code	Logo
1	ENGINEERING – INGEGNERIA INFORMATICA SPA	ENG	IT	
2	NATIONAL TECHNICAL UNIVERSITY OF ATHENS	NTUA	GR	
3	FUNDACION CARTIF	CARTIF	ES	
4	RHEINISCH-WESTFAELISCHE TECHNISCHE HOCHSCHULE AACHEN	RWTH	DE	
5	ACCADEMIA EUROPEA DI BOLZANO	EURAC	IT	
6	HOLISTIC IKE	HOLISTIC	GR	
7	COMSENSUS, KOMUNIKACIJE IN SENZORIKA, DOO	COMSENSUS	SL	
8	BLAGOVNO TRGOVINSKI CENTER DD	BTC	SL	
9	PRZEDSIEBIORSTWO ROBOT ELEWACYJNYCHFASADA SP ZOO	FASADA	PL	
10	MIASTO GDYNIA	GDYNIA	PL	
11	COOPERNICO - COOPERATIVA DE DESENVOLVIMENTO SUSTENTAVEL CRL	COOPERNICO	PT	
12	ASM TERNI SPA	ASM	IT	
13	VEOLIA SERVICIOS LECAM SOCIEDAD ANONIMA UNIPERSONAL	VEOLIA	ES	
14	ICLEI EUROPEAN SECRETARIAT GMBH (ICLEI EUROPASEKRETARIAT GMBH)	ICLEI	DE	
15	ENTE PUBLICO REGIONAL DE LA ENERGIA DE CASTILLA Y LEON	EREN	ES	
16	VIDES INVESTICIJU FONDS SIA	LEIF	LV	
17	COMITE EUROPEEN DE COORDINATION DE L'HABITAT SOCIAL AISBL	HOUSING EUROPE	BE	
18	SEVEN, THE ENERGY EFFICIENCY CENTER Z.U.	SEVEN	CZ	

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

Contents

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

Figures

Tables

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

Abbreviation and Acronyms

Acronym	Description
ALS	Airborne Laser Scanning
AI	Artificial Intelligence
API	Application Programme Interface
BBDVA	Building Big Data Value Association
BBDVC	Building Big Data Value Chain
BDVC	Big Data Value Chain
BEMS	Building Energy Management System
BIM	Building Information Modelling
CO_2	Carbon dioxide
DL	Deep Learning
DH	District Heating
DHN	District Heating Network
DSO	Distribution System Operator
DSM	Data Storage Management
ECM	Energy Conservative Measure
EEM	Energy Efficiency Measure
EE	Energy Efficiency
EP	Energy Performance
EPBD	Energy Performance of Buildings Directive
EPC	Energy Performance Certificate
ESCO	Energy Services Company
EU	European Union

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

Acronym	Description
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
ICT	Information and Communication Technologies
IEQ	Indoor Environmental Quality
IoT	Internet of Things
IPR	Intellectual Property Rights
KPI	Key Performance Indicator
LSP	Large Scale Pilot
M&V	Measurement and Verification
ML	Machine Learning
NIS	Network Information Security
PMB	Project Management Board
O&M	Operations & Maintenance
RES	Renewable Energy Sources
SECAP	Sustainable Energy and Climate Action Plan
SME	Small and medium-sized enterprises
WP	Work Package

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

Executive Summary

This deliverable entitled "*MATRYCS Technical and Security Specification*" reports the preliminary efforts and outcomes of Task 2.3 - Analytics Building Services Specifications - and Task 2.4 - Regulatory Framework for Data Protection, IPR and Ethical Issues. The scope of this deliverable is to deliver the technical and security specifications of MATRYCS framework, in order to provide useful information for the development of the first version of the MATRYCS reference architecture.

The work reported in D2.2 provides a high-level description of the expected specification and functionalities of MATRYCS framework and will feed the MATRYCS reference architecture, in order to guide relevant activities.

The in-detail definition of specifications is based on the Big Data Value Chain (BDVC) functionalities. They key activities of the BDVC are analysed, so as to set the basis for the technical and security specification elicitation. Therefore, this deliverable describes the methodology followed for the specification elicitation and analysis as well as their derivation from use cases and user requirements. The technical specifications have been derived based on the outcomes of the use cases as described in D2.1 - State-of-the-art analysis and Big Data Value Chain. Finally, a set of 105 user requirements were presented which led in 60 technical specification in total which were mapped to the three MATRYCS framework layers (GOVERNANCE, PROCESSING, ANALYTICS).

Finally, the technical specifications would be incomplete without including a detailed analysis on the regulatory framework for data protection, IPR and ethical issues. This analysis is focused on the identification of the security specification related to security and privacy of data taking into account possible ethical issues.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

1. Introduction

1.1 Purpose of the Deliverable

The scope of D2.2 "MATRYCS Technical and Security Specification" is to report on the progress of MATRYCS through the activities performed under Task 2.3 - Analytics Building Services Specifications - and Task 2.4 - Regulatory Framework for Data Protection, IPR and Ethical Issues - under which a set of technical requirements as well as the security requirements for the MATRYCS framework are derived.

The purpose of this deliverable is the definition and specification of the MATRYCS framework functionalities. A detailed analysis of the key stakeholders' requirements is conducted resulting in a unified Big Data Value Chain (BDVC) model. Additionally, this deliverable focuses on security and data privacy issues, analysing available standards about data protection in order to guarantee the security and protection of data collected by the Large-Scale Pilots (LSPs) during both the data acquisition phase and the service development.

Finally, legislative requirements and ethical issues are taken into consideration in order to assure that the MATRYCS framework will be compliant with legislative frameworks in both national and European levels.

This deliverable exploits the collected sets of user stories and use case scenarios as derived from actions under Task 2.2 User Stories and Requirements Analysis. The reported user stories and use case scenarios, which aim to include all stakeholders' needs and points of view, have been identified and reported in D2.1 - State-of-the-art analysis and Big Data Value Chain. This deliverable is prepared in accordance with the user stories collected in the abovementioned task and aspires to transform them into technical requirements.

1.2 Positioning within the Project

Deliverable D2.2 is the outcome of the efforts undertaken within the context of Task 2.3 - Analytics Building Services Specifications and Task 2.4-Regulatory Framework for Data Protection, IPR and Ethical Issues. D2.2 builds upon the outcomes of the efforts carried out within the context of WP2 System requirements and specifications until M6. D2.2 receives as input from D2.1 the first set of user stories and use case scenarios, as described by the targeted stakeholders.

Moreover, the results of this deliverable serve as input for Task 2.5 - Big Data, AI and IoT Reference Architecture and Alignment with Existing Frameworks and Architectures. The technical and security specification presented in this deliverable will set the basis of the MATRYCS framework reference architecture. Additionally, they will be considered as relevant information and inputs for WP3, WP4 and WP5 for the development of the MATRYCS framework components.

1.3 Deliverable Structure

The structure of the document is as follows:

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

- In Section 2, the BDVC is introduced, providing definitions of Big Data and BDVC and analysing the key high-level activities. A specific mention to BDVC for the buildings section is made. This section concludes with the BDVC of MATRYCS framework.
- In Section 3, the MATRYCS framework requirements are provided. An introduction to the MATRYCS framework is made. Moreover, the platform users are grouped to enable a more efficient presentation of the MATRYCS platform requirements. Finally, the user requirements are classified into functional requirements and non-functional requirements and data acquisition information is presented.
- In Section 4, the extracted technical requirements are presented. The technical requirements are grouped based on the platform modules and for each requirement the relevant layer of the MATRYCS framework and the respective user requirements are identified.
- In Section 5, the security specifications are analysed and grouped into three distinct categories: IPR and Ethical Issues Requirements, Network and Information Security Requirements and GDPR Requirements
- In Section 6, the future steps are discussed.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

2 Big Data Value Chain for MATRYCS Framework

2.1 Introduction to Big Data Value Chain

The Big Data Value Chain (BDVC) is introduced to describe the information flow within a Big Data system as a series of steps needed to generate value and useful insights from data [1]. The Value Chain enables the analysis of Big Data technologies for each step within the chain.

The emergence of a new wave of data from sources, such as the Internet of Things (IoT), Sensor Networks, Open Data on the Web, data from mobile applications and social network data, together with the natural growth of datasets inside organisations [2], creates a demand for new data management strategies which can cope with these new scales of data environments. Big Data is an emerging field where innovative technology offers new ways to reuse information and extract value from it. The ability to effectively manage information and extract knowledge is now seen as a key competitive advantage. To this end, significant number of organisations are building their core business on their ability to collect and analyse information, so as to extract business knowledge and insights. Big Data adoption is becoming an imperative need to gain competitive advantage.

Big Data definition implies that is the data with "high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision-making, insight discovery and process optimisation" [3]. Adam Jacobs describes Big Data as "data whose size forces us to look beyond the tried-and-true methods that are prevalent at that time" [4]. Big Data can be defined using the famous **5 Vs**: **Volume, Velocity, Variety, Veracity and Value** (Figure 1). **Volume** refers to the amount of data generated through websites, portals and online applications. **Velocity** refers to the speed which data are being generated with. **Variety** in Big Data refers to all the structured and unstructured data that has the possibility of getting generated either by humans or by machines. **Veracity** refers to the quality of the context of the data which will be utilised. Finally, **Value** is the ability to transform the volume of data into business value.

Volume	Velocity	Variety	Veracity	Value
• Amount of generated Data • Different units (KB,TB) • Saved in records, tables, files	• Speed of data generation • Generated in real time • In streams, batch, bits	• Structured or unstructured • Human generated or machine generated • Online and offline	• Accuracy and truthfulness of data • Data source, type, pre-processing • Interpretation and semantics	• Business value created by data • Social, economic, environmental value

Figure 1: The main attributes of Big Data

Within the field of business management, Value Chains have been used as a decision support tool to model the chain of activities, in order to deliver a valuable product or service to the market [5]. The Value Chain categorises the generic value-adding activities allowing them to be understood and optimised. A Value Chain consists of a series of subsystems, each of them comprising input, transformation processes,

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

and outputs. The European Commission (EC) sees the data Value Chain as the "centre of the future knowledge economy, bringing the opportunities of the digital developments to the more traditional sectors (e.g., transport, financial services, health, manufacturing, retail)" [6].

The Big Data Value Chain is the most established standard that includes a high-level modelling of all the required actions in a modern information system. The BDVC identifies the 5 key high-level activities (Figure 2) which are analysed bellow according to E. Curry [1]:

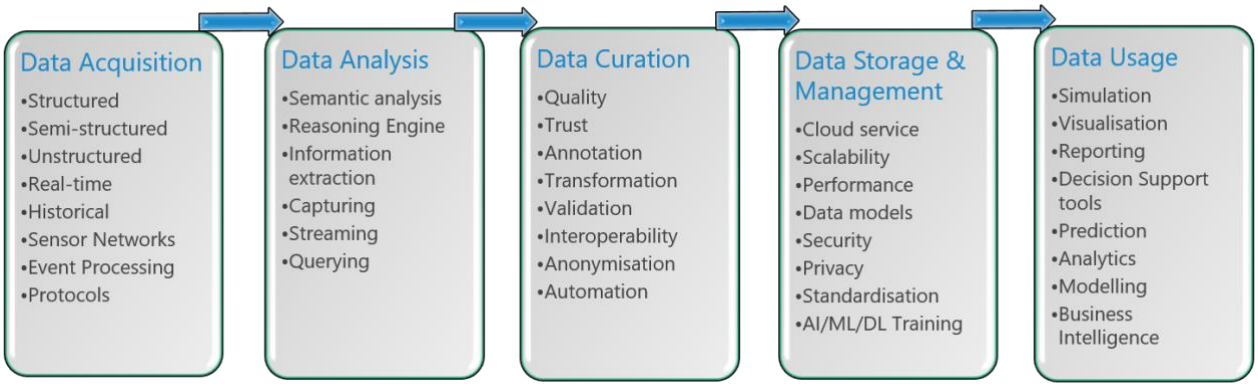


Figure 2: The key activities of the Big Data Value Chain

Data Acquisition incorporates all the necessary steps for collecting, filtering, and cleaning data. This procedure precedes the storage of data in a data warehouse or any other storage solution, as well as the data analysis process. Data acquisition is considered a very demanding and challenging process in terms of infrastructure requirements, as the infrastructure required to support the acquisition of Big Data must provide low, predictable latency in both data reception and query execution. The key challenges of this field include the ability to support flexible and dynamic data structures, in order to handle high transaction volumes, which take place in a distributed environment.

Data Analysis is the field responsible to convert primary data collected from different sources into a format that makes them available for specific uses and suitable to generate new information and to support decision-making. Data analysis includes the investigation, transformation and modelling of data with the aim of identifying relevant data, composing and extracting useful hidden information with high business potential. It is a wide field that combines knowledge from numerous sciences and utilises a range of tools such as data mining, machine learning and business intelligence.

Data Curation is the process of active data management throughout their life cycle, with the aim to ensure that data quality requirements are met, and that data can be used in a safe and efficient way. Data curation processes can be divided into the following sub-activities: content creation, selection, sorting, transformation, validation, and retention. All the above procedures are carried out by special curators responsible for completing all necessary steps, in order to improve the accessibility and quality of the data. Data curators, who are also known as data annotators, are responsible of ensuring that data satisfy specific properties such as trustworthiness, discoverability, accessibility, reusability etc.

Data Storage includes the continuous retention and management of data in a scalable manner to meet the need for direct access to data from a variety of applications. In recent decades, the dominant solution as a data storage standard has been Relational Database Management Systems (RDBMS) [7]. However, the ACID (Atomicity, Consistency, Isolation, and Durability) properties that guarantee database

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

transactions lack flexibility regarding schema changes and the performance and fault tolerance when data volumes and complexity grow, making them unsuitable for Big Data scenarios. Therefore, NoSQL technologies have emerged which present a wide range of solutions based on alternative data models achieving high levels of scalability.

Data Usage includes all actions that require access to data, their analysis, and all the tools and activities required to complete data analysis within the business. Data usage is necessary in the business decision-making process, as it can increase competitiveness through increased added value, cost reduction, or any other parameter that can provide an objective competitive advantage.

2.2 Big Data Value Chain for Building Sector

Big Data can have a vast impact on the building sector due to the amount of data generated by building and facilities. Therefore, measuring, storing and analysing all this data can be beneficial for their improvement and optimisation.

In terms of **increase of energy efficiency and building performance, as well as intelligent energy management of the built environment**, Big Data can play an important role, matching the energy demand to the production and reducing that demand as much as possible [8]:

If a large amount of historical data is available from different sources of the installations and buildings, such as energy consumption and production and weather data, it will be possible to develop accurate models of the energy production facilities and/or buildings. This, combined with real time data, would enable the energy production to be more closely aligned with the energy demand. Furthermore, if accurate weather forecasts are performed it would be possible to predict energy production if correlated with the weather and buildings demand.

In addition, it will be possible to improve Operations and Maintenance (O&M) by being aware of the normal operation of the installation or building from historical data. This will improve the performance of the installation and will also reduce the cost of the installations' maintenance.

If measurements of the indoor conditions of buildings are available, it is also possible to focus production on achieving the highest user's comfort for as long as possible.

Depending on the data available it is also possible to focus on different fields of the installations and/or buildings covering wide areas that can be optimised.

Related to the **support of building design**, Big Data potential is also high since, thanks to general or generic models that can be generated, it is possible to facilitate decision making and reduce design time. Achieving the most efficient design for the installation at the same time as reducing the effort as much as possible.

In relation to the **detection of funding opportunities in the building sector**, the improvement in the predictions that can be made would significantly help to develop more accurate Energy Performance (EP) Contracts due to the increase of information available. This will also enhance the de-risk of the different investment boosting the generation of more EP Contracts.

Buildings rarely perform as intended. It is well known that buildings have a high impact on the overall energy consumption and CO_2 emissions. Final energy consumption in building sector in Europe has increased by around 1.5% per year by non-residential building and by 0.6% for household since 1990

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

[9]. This problem combined with the greater awareness of the environmental impact of buildings and the concerns about energy availability, is remarking the important role of energy efficient technologies and effective delivery processes.

Indeed, the building sector is one of the largest consumers of energy in the world, with an estimated consumption of 40% of the global energy usage and one third of the global greenhouse gas emission. It is evident how the construction sector has a key role in effective climate policy [10], as well as the research in building energy efficiency is fundamental for reducing this value [11].

According to that, question arises spontaneously: Is there a reliable method to overcome this problem?

H. James Harrington said: *"Measurements is the first step that leads to control and eventually to improvement. If you cannot measure something, you cannot understand it. If you cannot understand it, you can*not control it. If you cannot control it, you cannot improve it". The meaning of this sentence, applied on buildings flows naturally on the continuous evaluation of building performance. The latter is achieved through detailed knowledge of the building and its proper management. In order to achieve this goal**, it is necessary to monitor the real behaviour of the building system through the use of Information and Communications Technology (ICT)** [12].

With the continuous development of sensor and actuator technology, network systems, cloud computing, (all of them elements of IoT systems), and software (Building Information Modelling - BIM) the volume of data coming from these systems is growing by the day with increasingly complex structure and forms [13].

The type of data generated on the building fields can come from different sources [14]:

- Data from design and construction (BIM)
- Utilities, building services, meters, building and technical management systems
- ICT systems and equipment
- Maintenance and operation of systems
- Enterprise systems (i.e. working schedule, occupancy data, performance reporting)
- Operational cost monitoring
- Post occupancy evaluation

Data from these sources can be used for different aims as shown in Table 1. The latter is a summary of the analysis conducted in Deliverable D2.1 "State-of-the-art analysis and Big Data Value Chain", by interviewing individual Large-Scale Pilots (LSPs) covering a wide range of application in the construction field.

Table 1: Summary of analysis conducted in deliverable D2.1 of MATRYCS regarding applications and cases related to increasing performance in buildings

LSP	NAME	Type of scale	How Big Data can improve the performance of the pilots reducing the energy consumption?
1	BTC	BUILDING	Action Plans to identify energy efficiency measures
			Definition of potential energy efficiency and flexibility projects that can implemented through energy performance contracting.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

LSP	NAME	Type of scale	How Big Data can improve the performance of the pilots reducing the energy consumption?
2	FASADA	BUILDING	Developing BIM models or digital twin of the building to simulate building performance
			Definition of possible scenarios to be applied aiming the reduction of energy consumption of the building
3	VEOLIA	BUILDING	Building energy consumption prediction
4	ASM	BUILDING	Building management optimisation: improve the capability to deploy optimal comfort-aware building energy consumption management services
			Operational efficiency: the local smart distribution grid needs to be coordinated with the building consumption
			Decentralised grid management: optimisation of building management to exploit Renewable Energy Sources (RES)
5	COOPERNICO	BUILDING	Increase efficiency of current energy assets
			Identity energy efficiency opportunities
			Collaborate with local entities to foster collective improvements in management of local energy assets (existing or new ones)
			Map local electricity consumption to identify citizens in energy poverty condition
6	VEOLIA	DISTRICT	District Heating Network (DHN) operation and optimisation
			Measurement & Verification of energy savings after application of an Energy Conservation Measure (ECM)
			Implementation of energy conservation measures
7	ICLEI	CITY	Development of Sustainable Energy and Climate Action Plan (SECAP)
8	GDYNA	CITY	Analysis of communal buildings to understand their performances
9	EREN	REGIONAL	Supporting data share and harmonisation of Energy Performance Certificates (EPC)
			Facilitating compliance and checking of EPC
			De-risking investments with better Energy Performance Certificate proposals
			Calculating actual energy savings
10	LEIF	NATIONAL	Define list of the energy efficiency measures and their costs
			Calculate KPIs, energy performance, investment risk and cross-domain financial
11	HOUSING EUROPE	EUROPEAN	Select the optimal set of policy options (cost efficient, feasible, and ambitious at the same time) to include in the National energy and climate plans (NECPs) and Long-term renovation strategies
			Make informed decisions based on a set of feasible and data-tested options
			Make reliable predictions about outcomes

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

Based on the outcomes of the Table 1, the use of Big Data in the building sector can cover different purposes, such as:

- Understand and predict the energy behaviour of the building or district.
- Verify the real saving achievable after the application of an energy conservation measure using a Measurement and Verification (M&V) protocol.
- Simulate building performance (Digital Twins).
- Select the best options for policies.
- Define the best energy efficiency measures to by applied in a building to reduce the energy consumption and increase the Indoor Environmental Quality.

The application of the Building BDVC is fundamental to reach the outcomes defined in Table 1. Smart technologies enable the **acquisition**, **analysis**, **curation storage** and **usage** of data permitting not only observation, monitoring and control of individual processes in isolation, but analysis of how they interact or how a change in one affect another.

According to the different user, the use of the Building BDVA allows to:

1) Identify and benchmark the best policies implemented in a specific field, which can be replicated to achieve the maximum impact. The connection of these results with data related to financial and business models allow the user (**policy maker**) to have a better overview of the possible actions to finance.
2) Help **investors** to reduce the uncertainty linked to energy efficiency investments. To this end, it is important to store and process historical data, apply specific machine learning algorithms to better predict the energy consumption, calculate and monitor the energy saving achievable after the application of an Energy Conservation Measure (ECM). The use of Measurement and Verification approach based on the International Performance Measurement and Verification Protocol (IPMVP) is fundamental to have a reliable result.
3) Improve, harmonise and increase the reliability of the Energy Performance Certificate (EPC). The public awareness of the importance of EPC will be raised, giving **private and public investors** the possibility to better plan interventions.
4) Detect building/systems faults, improving the overall energy efficiency. Data coming from Building Management Systems will be elaborated in real-time ensuring the **energy/facility manager** to take control over its data and expose flexibility to various service providers. In this way, the overall operational efficiency has increased supporting the maintenance and planning activities.
5) Benchmark the energy performance of the building using both real and simulated data (i.e. Digital Twin). The **designer** can reduce the possible errors generated in the design phase, reducing the gap between the data coming from an energy model (simulated data) and data coming from building sensor.

2.3 Big Data Value Chain of MATRYCS framework

According to the BDVC for building sector as analysed in Section 2.2, MATRYCS framework is following

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

the key activities to ensure a successful implementation of MATRYCS ecosystem and reach outcomes defined in Table 1. Figure 3 presents the key activities to be taken into consideration in order to develop MATRYS framework along with the respective Work Package (WP) under which they are being implemented.

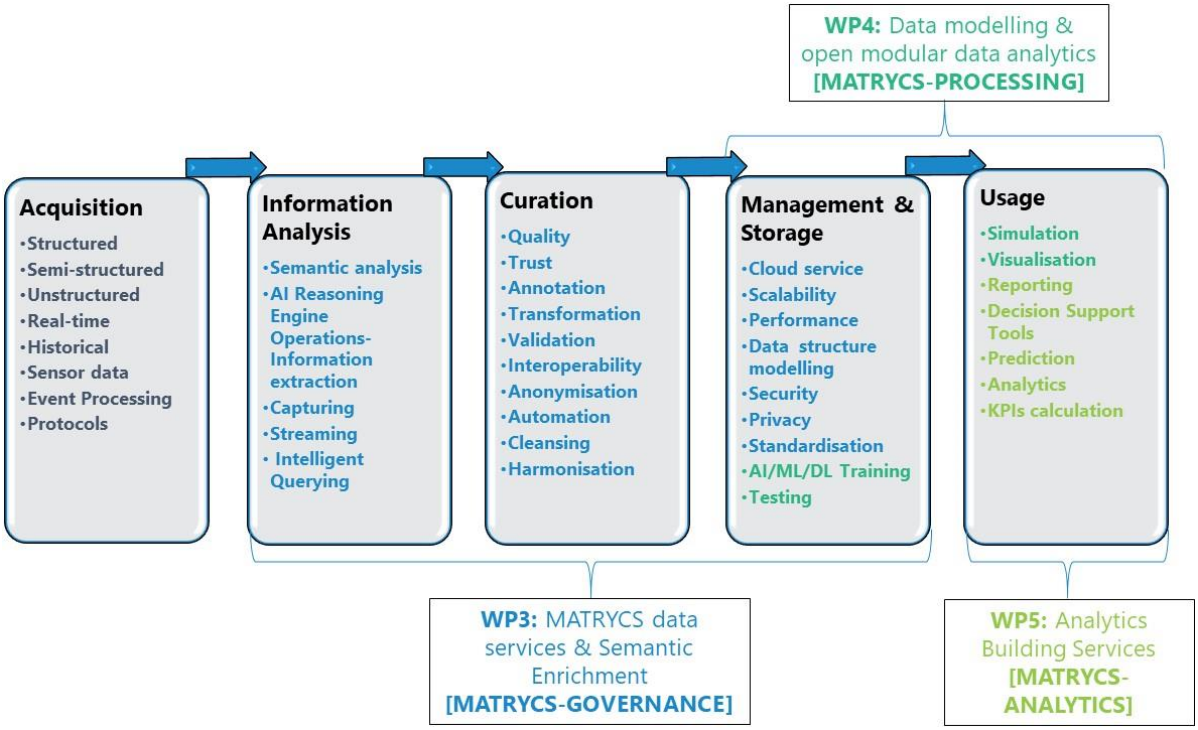


Figure 3: BDVC for MATRYCS framework

The BDVC of MATRYCS framework activities set the basis in order to define the technical specifications required for the three different levels of MATRYCS: MATRYCS-GOVERNANCE, MATRYCS-PROCESSING, MATRYCS-ANALYTICS.

In the following sections, the MATRYCS framework requirements analysis follows the key activities of the BDVC of MATRYCS framework. The scope is to cover the identification of all requirements per activity to ensure the value created from one level can be forwarded to the others through a mutual beneficiary cooperation.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

3 MATRYCS Framework Requirements

3.1 General Description

The main objective of MATRYCS is to address the challenges in Big Data management for buildings, creating a unified framework which aspires to be an open solution for stakeholders operating in this sector and to open new market opportunities. In this scope, MATRYCS aims at the definition and deployment of a **Reference Architecture for Buildings Data** exchange, management, and real-time processing, as well as at the encapsulation of this architecture to an **Open, Cloud-based Data Analytics Toolbox (MATRYCS Modular Toolbox)**.

Overall, the MATRYCS Modular Toolbox consists of three main pillars: i) the **MATRYCS-GOVERNANCE** including modules related to data collection, semantic annotation and distributed storage, ii) the **MATRYCS-PROCESSING** including ML and DL models and iii) the **MATRYCS-ANALYTICS** providing a set of analytics tools as a service (SaaS/PaaS/IaaS models).

The **MATRYCS-GOVERNANCE** (Figure 4) constitutes the first layer of data analysis right after their acquisition, acting as a link between data providers and data users. MATRYCS-GOVERNANCE is based on two main components: the first component concerns the guarantee of secure access to the acquired data through modern technologies that ensure direct access to the security, reliability and transparency of data. The second component focuses on data integration, efficient data pre-processing, semantic enrichment and data query mechanisms in an efficient way.

Interoperability Service Module	• APIs enabling communication between data from different sources • Development of open data models • Use of 3rd party static datasets
Data Cleansing, Curation, Assess Policy and Anonymization Module	• Code Snippets for Data Pre-processing • Real-time Data management continuous streams • Data Quality analysis, Security and GDPR Issues
Data Capturing and Streaming Module	• API for Data Collection • Streaming Data Mechanism
Data Semantic Enrichment Module	• Construction of Common Data Model • DBMS design • Big data storage
Reasoning Engine	• Development of state-of-the-art Reasoning Engine • Processing in real time with continuous flow • Incorporation of Big volume of data
High Performance Distributed Query Engine	• Big Data Query Engine like Presto and Tez • High Performance requirement

Figure 4: MATRYCS-GOVERNANCE Components

The **MATRYCS-PROCESSING** (Figure 5) includes a library of reusable AI-based ML and DL models that will be developed in order to provide quick adaptation and reusability of machine learning models along

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

different use cases within pilots. This module will encapsulate the AI components of the MATRYCS framework, serving as an intermediate layer between MATRYCS-GOVERNANCE and MATRYCS-ANALYTICS, as it exploits data and modules of the MATRYCS-GOVERNANCE in order to perform intelligent algorithms and generate innovative, reusable models.

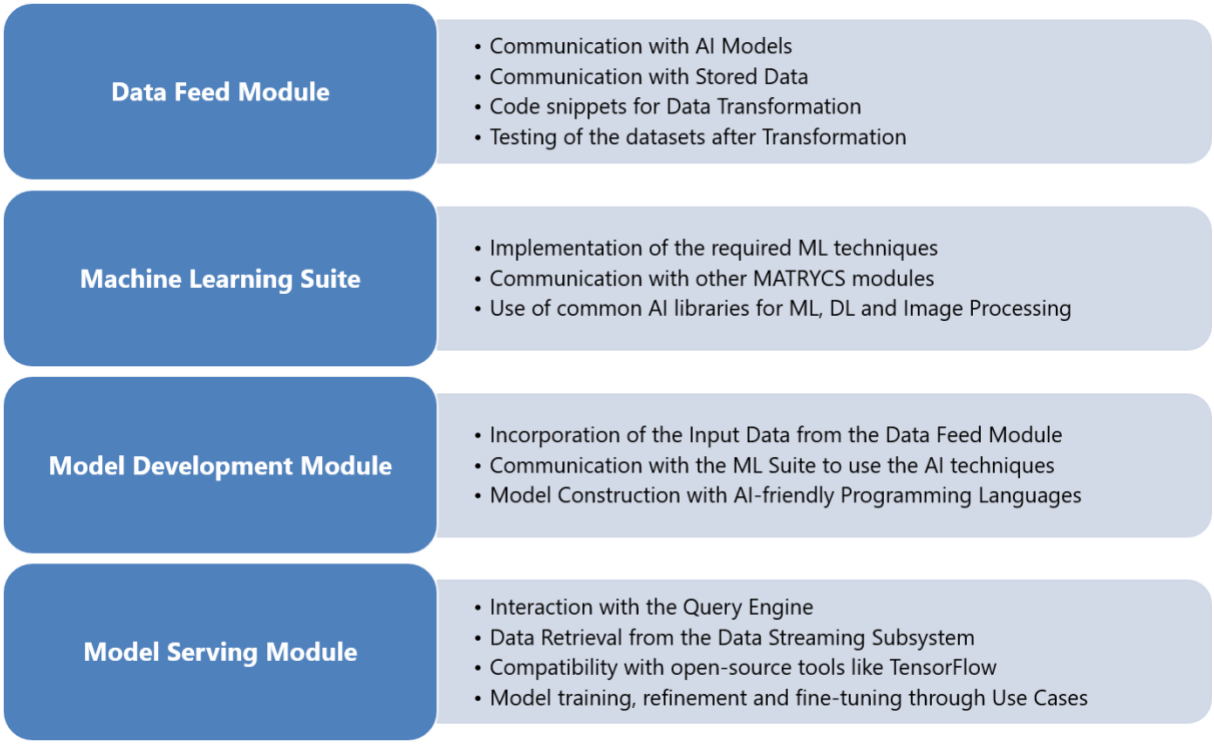


Figure 5: MATRYCS-PROCESSING Components

The aim of the **MATRYCS-ANALYTICS** is to create a set of analytics tools as a service. The services can be mainly divided in Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) level.

The **Software as a Service (SaaS)** will develop services based on LSPs input and requests. The services can be divided according to the different application fields, such as building performance and Indoor Environmental Quality (IEQ), building infrastructure, policy impact assessment and building efficiency investments (Figure 6).

The **Platform as a Service (PaaS)** will offer a "virtual workbench", to provide developers and stakeholders an environment that will help in building new services based on distributed data and AI models. It will incorporate and provide access to different MATRYCS assets and tools including data and a series of trained ML/DL models. The virtual workbench platform will provide users an intuitive and easy to use Graphical User Interface (GUI) to create new services in the building sector.

The platform will integrate two modules:

Building Data & Services Innovation Hub: It will provide SMEs, developers and potential innovators a graphical interface that will facilitate the development and deployment of new applications on the base of the existing data sets and the available set of the developed and trained models being produced by MATRYCS.

Data Analytics Environment: It will allow users (e.g., analysts) to perform freeform queries and data analytics on the assets which are accessible to the platform (models and data).

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

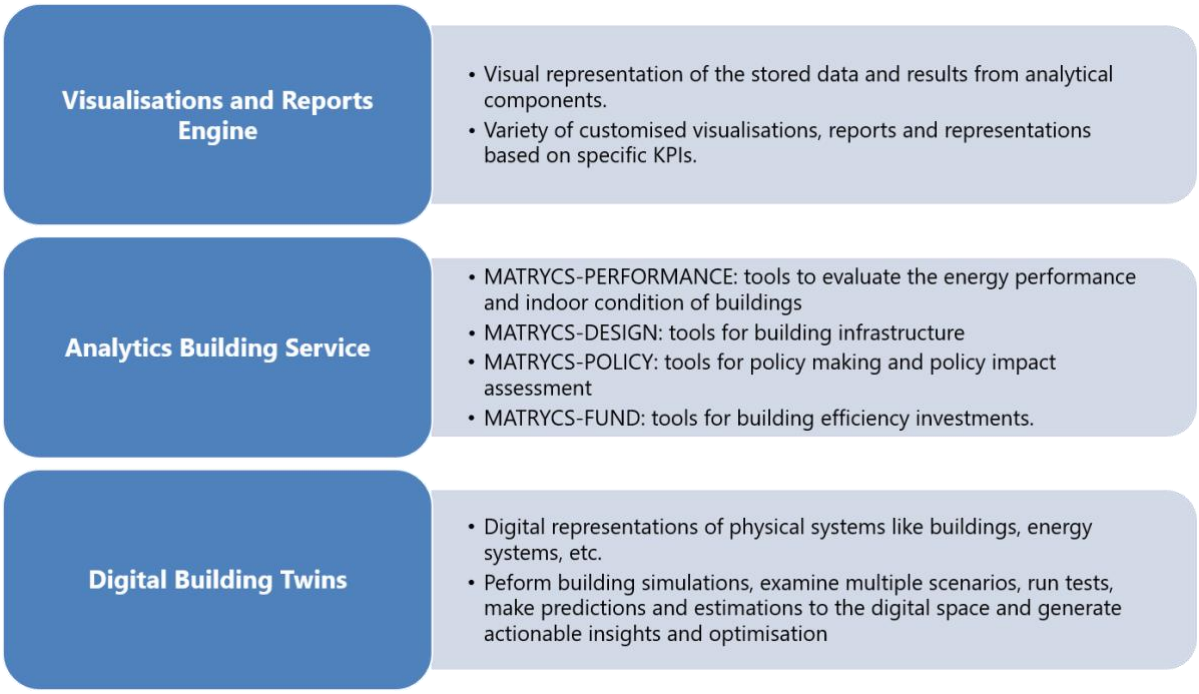


Figure 6: MATRYCS - ANALYTICS Software as a Service (SaaS)

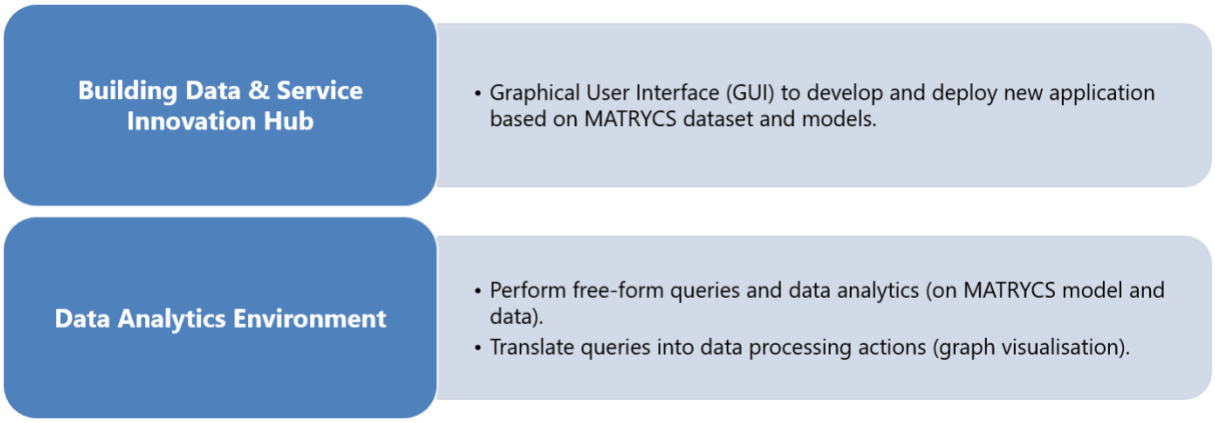


Figure 7: MATRYCS - ANALYTICS Platform as a Service (PaaS)

At **Infrastructure as a Service (IaaS)** level, the platform will package its services, as an adapted open - source sandbox, in container-based workflows that could be flexibly deployed over cloud infrastructures through a service orchestrator. The pre-existing available local-level capabilities for sensing, gathering, integrating and processing data (e.g., IoT devices) through available in-house analytics platforms can be augmented with the capabilities of the proposed analytics reference toolbox.

3.2 Data acquisition

3.2.1 Large Scale Pilots

As data acquisition incorporates all the necessary steps for collecting to be used, within this section information about the process of data acquisition from each LSP is presented for the needs of the implementation of the MATRYCS functionalities and services. The scope is to understand the datasets,

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

their flexibility and structures and ownership and security aspects. More specifically, the following characteristics per LSP data sets are specified:

- Description of datasets
- Dataset volume (GB/day or GB/year)
- Starting date of data collection
- Current data storage format and communication interfaces
- Available Databases and data ownership
- Reliability and security of datasets

Before presenting the data acquisition information per LSP, it should be highlighted that the following information is in preliminary stage according to the LSPs description of the datasets that they will provide. In the following iterations of the services' implementation the following information may be enriched.

BTC (LSP1) has data from various sources, such as electricity, heating, solar plant production etc. In a yearly basis, the volume of each data category will not be more than 100MB, while the time horizon depends on the datasets (some datasets have a 10-year time horizon beginning from 2010, while other datasets are more recent beginning from 2017). All datasets are made through SQL queries and the current format of data is MSSQL (except from Waterpark Bracelets data which are stored in Oracle DB). A SQL database is available and all datasets are structured. The ownership of data is strictly private, supported by the need of login to the SQL database and a VPN to the owner's network.

FASADA (LSP2) provides both dynamic and static data. In the case of dynamic data, FASADA measures temperature, humidity and CO_2 inside buildings and also provides thermographic maps for two rooms. In the case of static data, it provides the (Building Information Modelling) BIM of a building, along with cloud points gathered with Airborne Laser Scanning (ALS) method. The starting year of the data gathered is 2018 and the data volume is 100Kb per hour. The data format varies from CSV and XLSX to JPEG and PNG for the visualisation of the thermographic maps. The data ownership belongs to FASADA and the datasets are gathered by UNIVPM in the context of P2ENDURE project.

VEOLIA (LSP3 and LSP6) measures and stores data from different buildings and facilities, therefore the volume of data depends on the number of monitored variables and the storage frequency. The creation of the csv file by Veolia is made through an SQL query made to the SQL database of Veolia. This database is constantly synchronising via VPN Internet network the Veolia SQL database and the local SQL database. This local SQL database is stored in the 963 Trend Control Systems® supervisor software installed in Veolia's local server in each facility. This server gathers the data from the controllers through a LAN TCP/IP network every day.

ASM (LSP4) gathers static and semi-static data concerning Building Energy Management Systems (BEMS), weather information and data from smart meters inside buildings, among others. The current storage format of data is JSON, CSV files and reports, while the communication interfaces provide a REST API and http protocols. Data security is ensured through VPN connections and user credentials. The data volume of each category is less than 1 MB per day and the starting year of data gathered is 2018 (except from weather information which is 2014). Finally, the data ownership is private.

COOPERNICO (LSP5) gathers dynamic, static and semi-static data concerning PV production, billing information, smart meter data, residential energy data, social information and weather data (through an

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

API). The storage of the data is in CSV files and possible connection with an SQL database. An API for accessing smart meters data is available through a python library, as well as a monitoring platform that gathers near real-time production data of their solar plants but with no communication interface available as of now. Data volume for all files is considered to be in terms of a few GB per day. Velocity varies from 15 mins stream up to a day for non-static data. Finally, the data ownership is private either from COOPERNICO's side or their clients.

ICLEI's (LSP7) data concern various categories, such as city demographic data, climate actions, city-wide GHG emissions etc. Data are communicated via a File Transfer Protocol (FTP) and the current storage format of the provided datasets is CSV. A database is not currently available but is under development with MySQL technology. The communicated data are open and not subject to ownership regulations, while the security of transmissions is supported with a restricted IP address.

GDYNIA (LSP8) datasets concern building consumption data such as electricity consumption, water/gas usage per month and building characteristics (year of construction, insulation etc. Possible further datasets are cadastral data, EPC registers, thermovision images of buildings, invoices and energy prices in area. Data are possibly static, in various formats such as CSV, PDF, XML and JPG. The proposed volume is low, around 2GB/year for non-static data. No communication interface for these data is required.

EREN (LSP9) datasets consist of regional Energy Performance Certificates of different regions, cadaster data, weather data (historical and forecasting). The files are in CSV, JSON, XML, geoJSON and GML format with https, ATOM Service (sml) and dedicated APIs as interface, for the different communication with the datasets. There are also public databases to access for some of these data. The velocity varies from hourly to daily frequency for non-static data, with the volume be at 50-60 mb per day. The privacy of the data is open source, except for the Regional EPC repositories which are owned from the public administration authorities.

LEIF (LSP10) stores both dynamic and static data. The main categories of dynamic data are temperature and humidity data through sensors and electricity data through an electronic data monitoring system. On the other hand, the categories of static data are energy consumption of various activities like water heating, house heating, electricity etc. The starting year of data collection is 2013 and the main communication interface is the http protocol and local network. Data are stored in CSV, XLSX and PNG format and the ownership of data belongs to LEIF.

HOUSING EUROPE (LSP11) datasets consist mainly of open European datasets, such as governmental and EU information that are based on reporting, statistics and EU projects. For these datasets, public access is available for most of them through their dedicated web interface. The data are offered in CSV, JSON, XML and TXT files. The data are static as they are official stats and numbers, and their privacy is public.

3.2.2 Other repositories

Aside from LSPs' data, **open data sources** will be integrated in different manners, taking into account if they are more related with the data model or if they are related with the service to be developed. More information on datasets and repositories that at this stage of the project have been identified as relevant for MATRYCS are provided by D2.1 in Chapter 4, and more specifically Table 27 which is to be regularly expanded and updated throughout the project duration.

For example, in LSP9, data from cadastre for the Castilla y León region will be used. The idea is to

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

integrate this data in the MATRYCS data repository, since it could be used as the geometrical basis for the digital twin in this pilot. Other datasets that will be used come from TABULA/EPISCOPE or other similar sources. These datasets contain information of the different typologies that will be processed to the service in order to extract the relevant information for the building in the region.

In the case of the Spanish cadastre a script has been developed in order to download the information of the buildings for the region of interest. At this respect 2248 GML files (each file corresponds to a municipality in the region) containing information about buildings and 2248 GML files containing information about buildings parts (more related with the geometry of the different parts of the buildings with the aim to generate complex geometries) have been downloaded. It is important to note that the update rate of the published cadastre data is very low, so it is not expected to have a new version of the cadastre during the project. There are in total 13.2 GB of data.

In the case of the information for the typologies, this data will be used directly in the service and will be considered as configuration information for the service. An update on the use of data from external sources will be reported in D2.4 MATRYCS Reference Architecture for Buildings Data (final version), once the reference architecture is finalised (M18).

3.3 Users' Overview

MATRYCS envisions to benefit several stakeholders, therefore, their contribution to the MATRYCS framework developments is significant. At an early stage, stakeholders that derived after discussion with LSPs leaders participated in order to provide useful insights and guidelines for the development of targeted cases based on their requirements. These stakeholders who are expected to be benefited from the outcome of this project are identified through the 11 LSPs, as defined in D2.1 State-of-the-art analysis and Big Data Value Chain. This section provides a short overview of the LSPs and the respective stakeholders. However, more information can be found in D2.1. Therefore, Task 2.2 defines the stakeholders' approach as a twofold process: first, by identifying the Target Groups that apply to each LSP, and then, by assigning different Personas to each of the Target Groups defined.

A **Target Group**, as defined in Deliverable 2.1, is the group of potential users of the service(s) that is going to be provided by the pilots. They are the stakeholders or other kind of users/beneficiaries from the service. Target Groups have been categorised in order to have a common classification and naming that has been refined when pilots provided their inputs. Table 2 below includes the Target Groups along with their definitions.

Table 2: MATRYCS Target Groups categorisation

Target Group	Brief definition
Researchers	Research organisations and universities
Facility Managers	Manager in charge of the correct functioning of their facilities
ESCOs	Energy Services Companies
Real estate developers and managers	Developers and managers of real estate

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

Target Group	Brief definition
Utilities	Public and private utilities and aggregators
Institutions	Relevant national and European institutions or departments
Investors	Industry and investors related to the real estate management and development
Policy makers	Policy authorities and policy makers (local to European level), such as urban planning department
Designers	Architects and engineers in charge of the design, retrofitting or decommissioning of buildings.
Constructors and contractors	Construction companies and consulting firms
Citizens, tenants and Owners	Citizens and individual owners
SMEs, Companies and entities	Small and Medium enterprises, related and not related to construction

Personas are fictional characters created to represent the needs, desires and behaviours of the Target Groups to ensure that we are thinking from their perspective. Each Persona captures a different point of view of each Target Group. Thus, for each Target Group, several Personas or characters can exist. Table 3 collects the complete recording of the Personas of each pilot, related to the Target Groups to which they are associated for their creation.

Table 3: MATRYCS Pilots and Personas (related to their Target Group)

LSP	Target Group	Persona
LSP1: BUILDING OPERATION [BTC]	Facility Managers	Informed facility manager
		Sustainability supporting facility manager
		Cost-conscious facility manager
	Citizens, tenants and owners	Informed owner
		Sustainability supporting owner
	Utilities	Informed utility
		DSM aware utility
	ESCOs	Informed ESCO

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

LSP	Target Group	Persona
		Sustainability supporting ESCO
LSP2: BUILDING REFURBISHMENT [FASADA]	Citizens, tenants and owners	Personnel of the kindergarten
		Management staff of educational buildings in the municipality (City of Gdynia)
		Users of the kindergarten
	Designers	Designers willing to plan efficient renovation activities
LSP3: ESCO SUPPORT SERVICES [VEOLIA]	ESCOs	Advanced ESCO
		EPC facilitator
	Citizens, tenants and owners	Motivated owner
		Careless owner
		Disbelieving owner
	Investors	Benefit-oriented bank
		Sustainability supporter bank
LSP4: SMART BUILDING AS ACTIVE NODE OF A SMART GRID [ASM]	Citizens, tenants and owners	Sensitive office employee
		Informed office employee
		Reluctant office employee
		Dynamic prosumer
		Static prosumer
	Utilities	Large scale DSO head
		Small scale DSO head
		Public sector energy manager
	ESCOs	Private sector energy manager

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

LSP	Target Group	Persona
LSP5: ENERGY COMMUNITIES [COOPERNICO]	Citizens, tenants and owners	Prosumer
		Environmentally conscious owner
	Policy makers	Local authority
	SMEs, Companies and entities	Eco-friendly SME
LSP6: DISTRICT HEATING NETWORK [VEOLIA]	Facility Managers	Sustainability supporter DHN manager
		Cost-conscious DHN manager
	Designers	Profit-oriented designer
		Efficiency-oriented designer
	Investors	Informed investor
	Citizens, tenants and owners	Informed owner
LSP7: SECAPS [ICLEI]	Institutions	Urban planning professional
		Local transport department director
	Policy makers	City council member
		European commission staff
LSP8: ONE-STOP-SHOPS [GDYNIA]	Citizens, tenants and owners	Real estate owner
		Building administrators and managers
	SMEs, Companies and entities	Renovation contractors and auditors
	Institutions	Workers of Gdynia and other municipalities
LSP9: ENERGY PERFORMANCE CERTIFICATES	Designers	EPC issuers
	Policy makers	EPC's Regional Registry Management Service

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

LSP	Target Group	Persona
[EREN]		Regional energy efficiency strategy planner
	Citizens, tenants and owners	Buildings' buyers and sellers, landlords and tenants
LSP10: DE-RISKING EE INVESTMENTS [LEIF]	Institutions	Ambitious head of department
		Project manager
	Facility Managers	Responsible manager
		Careless manager
	Policy makers	Informed policy maker
	SMEs, Companies and entities	Director of finance
LSP11: POLICY MAKING AND IMPACT ASSESSMENT [HOUSING EUROPE]	Policy makers	Member of the Parliament (Energy and Housing Committees)
		Senior official in the Commission for Europeans
	Real estate developers and managers	Director of Project Development among one of the biggest housing managers
		Manager of a small local housing cooperative
	Citizens, tenants and owners	Old citizen in rented house
		Student in community group to convince of installing solar panels in the roofs
	Constructors and contractors	Worker of a large construction company
		CEO of a large manufacturer of energy systems

3.4 MATRYCS Use Cases

A use case provides a complete and systematic description of the way that users will interact with the MATRYCS platform as well as the individual tasks which they need to complete. It, also, records the system's behaviour to any request from the user's point of view. Each use case is expressed as a sequence

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

of well-described steps, beginning with the user's target and ending with the fulfilment of this target.

Use cases are essential for any computational system -or generally any other system- as they include all the targets of the system along with their cost and their complexity, thus enabling an estimation of the total complexity of the system. An additional advantage of analysing the use cases is that they provide information about the potential dangers and obstacles that might occur in any subsystem.

The use case elicitation process is composed of several steps, demanding the cooperation of both the LSPs and MATRYCS technical partners. The use cases were based on specific usage scenarios which were developed according to the needs of the platform's personas. These personas filled in a detailed form, analysing their needs and how they could potentially benefit from the MATRYCS platform. Based on these forms, a series of user stories were extracted, leading to the specification of the abovementioned usage scenarios.

More specifically, each (LSP was asked to find all the personas that could benefit from the MATRYCS Toolbox, in order to have a complete record of the potential end-users. After this process, LSPs, with the assistance of the technical partners described the user story of each persona. Each user story, in its part, resulted in one or more usage scenarios, which represent the applicability that the platform could have for each persona. A detailed description of the use case extraction procedure is included in D2.1 State-of-the-art analysis and Big Data Value Chain.

Figure 8 describes the use cases extraction process and their position in the user requirements elicitation process.

Figure 8: MATRYCS Requirements Elicitation Process

A use case should be described in an easy-to-understand narrative. In the context of the MATRYCS, a complete framework has been developed for the specification of the use cases. This framework consists of four basic components which include different details about the use case. These components are described below:

- **Description:** The first component involves the general description of the use case. Firstly, each use case is provided by a unique Use Case ID, the Scale(s) and the Domain(s), as well as the name of the use case. Additionally, the use case's scope, objective(s) and related business case(s) are analysed. Finally, this component incorporates the narrative of the use case and the general remarks – if any.
- **Diagram(s) of Use Case:** The second component includes the use case diagrams. Use case diagrams consist of actors, use cases and their relationships. Such diagrams' importance lies to the fact that they model the behaviour of a subsystem of an application, because single use case diagram captures a particular functionality of the system, along with the interactions of the system and the actors. Therefore, in order to model the entire system and develop its architecture, a number of use case diagrams are used.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

- **Technical Details:** The third component concentrates all the technical specifications of the use case. This is a very significant step for a use case definition because it plays an important role for the mapping of the use cases to the users' requirements and the functional and non-functional requirements elicitation. Therefore, the reporting of the technical details is made as follows: The actor and the system of the use case are clearly defined, as well as the triggering event, the pre-conditions and the post-conditions, in order to provide a better understanding of the technical elements involved.
- **Step by Step Analysis of Use Case:** Finally, the fourth component presents a step-by-step analysis of the use case. More specifically, for each scenario involved in the use case, the actors are mentioned and a series of steps are recorded until the fulfilment of the use case. In each of these steps, the event is presented, the activities of the event are descripted in detail and finally the exchanged information is noted.

In the context of this deliverable, the most important information of each use case is presented. The following information is included:

- **Use Case ID**: A unique identification for each use case.
- **Name**: The name of the use case providing a short description of the task.
- **Objective(s)**: The main user targets that are satisfied by the use case.
- **Related Actors**: The involved users that will benefit.
- **Involved Services**: The services which connect with the use case and will be utilised to satisfy the usage scenarios.

Table 4 includes the 25 use cases of the MATRYCS project:

Table 4: Use Cases Short Description

ID	Use Case Name	Related LSPs	Involved Services	Related Actors
UC01_01	*Action plans for preventive maintenance*	*LSP1*	*s1.2, s1.4*	*Facility Managers*
UC01_02	*Action plans for performance improvements*	*LSP1, LSP4*	*s1.1, s1.2*	*Facility managers, Sustainability supporting owner, Informed owner of shops, Sensitive office employee*
UC02_01	*Building evaluation of ECM / Building refurbishment*	*LSP2*	*s2.2, s4.2*	*Management staff of buildings, designers*

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Use Case Name	Related LSPs	Involved Services	Related Actors
UC03_01	*M&V plan creation*	*LSP3, LSP6*	*s4.1*	*Advanced ESCO Informed owner*
UC03_02	*Savings report generation*	*LSP3, LSP6*	*s4.1*	*Advanced ESCO, Informed owner*
UC03_03	*Increased knowledge and confidence of the owners of benefits of EEM application*	*LSP3, LSP6*	*s1.3*	*Advanced ESCO, EPC facilitator, Owner, Bank, investor*
UC04_01	*Optimisation of electrical distribution system operation*	*LSP4*	*s1.1, s1.5*	*Small Scale DSO Head*
UC04_02	*Action plans for performance improvements*	*LSP1, LSP4*	*s1.1, s1.2*	*Facility managers, Sustainability supporting owner, Informed owner of shops, Sensitive office employee*
UC05_01	*Predictions about carbon footprint*	*LSP5*	*s1.1*	*EV owner, eco-friendly SME*
UC05_02	*Identification of PV performance issues*	*LSP5*	*s1.1, s1.4*	*Prosumer*
UC05_03	*Optimisation of self-consumption of PV system*	*LSP5*	*s1.1, s1.2*	*Prosumer*
UC05_04	*Creation of technology catalogues*	*LSP5*	*S2.1*	*Prosumer*
UC06_01	*M&V plan creation*	*LSP3, LSP6*	*s4.1*	*Advanced ESCO Informed owner*
UC06_02	*Savings report generation*	*LSP3, LSP6*	*s4.1*	*Advanced ESCO, Informed owner*

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Use Case Name	Related LSPs	Involved Services	Related Actors
UC06_03	*Increased knowledge and confidence of the owners of benefits of EEM application*	*LSP3, LSP6*	*s1.3*	*Advanced ESCO, EPC facilitator, Owner, Bank, investor*
UC06_04	*Operation of DH Network*	*LSP6*	*s1.1, s1.5*	*Technical Manager*
UC07_01	*Policy-makers decision-making support*	*LSP7*	*s3.1, s3.3*	*Urban Planning Professional, Local transport department director*
UC07_02	*Building's investment prioritisation*	*LSP7*	*s3.1, s3.3*	*City council member*
UC08_01	*Geo-municipality service*	*LSP8*	*s0.2, s1.3*	*Real state owner*
UC09_01	*EPCs data error checker and validator*	*LSP9*	*s3.2*	*Energy Performance Certification (EPC) issuers*
UC09_02	*Visualisation of estimated EPC*	*LSP9*	*s3.2, s1.3*	*Regional Energy Efficiency planner*
UC10_01	*Risk assessment of funding in energy efficiency investments*	*LSP10*	*s3.3*	*Informed policy maker*
UC11_01	*Portfolio optimisation*	*LSP11*	*s3.3, s4.2*	*National authority - policy maker*
UC00_01	*Digital Twin creation*	*LSP1, 2, 6, 8, 9*	*s0.1*	*DT creator; Service developer*
UC00_02	*Services integration in Digital Twin*	*LSP1, 2, 6, 8, 9*	*s0.1*	*DT creator; Service developer*

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

3.5 Functional and Non-Functional Objectives

In order to record users' requirements of the MATRYCS framework, a multi-level process in conjunction with the LSPs was required. Following the use case extraction procedure, final steps took place in order to extract functional and non-functional user requirements from the use cases. The user requirement elicitation process results in two categories of requirements which are described below:

Functional requirement is the declaration of the intended functionality of a system and its components as reported by a hypothetical non-technical observer [15]. The functional requirement is facilitating the development team to determine the expected behaviour or output of the system in the case of a certain input and in which a technical problem is addressed. Additionally, within the functional requirements the outputs of the envisioned product increment when receiving the described input is described.

Non-functional requirements define system attributes such as security, reliability, performance, maintainability, scalability and usability. Also known as system qualities, they are just as critical as functional requirement as they safeguard the usability and effectiveness of the entire system. Failing to meet any of them can result in systems failure to satisfy business, market or user needs. In some cases, non-functional requirements cannot be resolved to one function, component or layer in the architecture. And in other cases, they cannot be tested directly. Nevertheless, they must be kept in mind for choosing the right design and implementation of a system [16].

Before presenting the MATRYCS users' requirements, some elements of the elicitation procedure should be described. Therefore, the special characteristics of the requirements are presented [17], ensuring that MATRYCS requirements meet quality standards of a good requirement:

- **Unitarity/Cohesion:** The requirement addresses one and only one thing.
- **Completeness:** The requirement is fully stated in one place with no missing information.
- **Consistency:** The requirement does not contradict any other requirement and is fully consistent with all authoritative external documentation.
- **Atomicity:** The requirement is atomic and it does not contain conjunctions.
- **Traceability:** The requirement meets all or part of a business need as stated by stakeholders and authoritatively documented.
- **Unambiguousness:** The requirement is concisely stated without recourse to technical jargon, acronyms (unless defined elsewhere in the Requirements document), or other esoteric verbiage. It expresses objective facts, not subjective opinions. It is subject to one and only one interpretation. Vague subjects, adjectives, prepositions, verbs and subjective phrases are avoided. Negative statements and compound statements are avoided.
- **Verifiability:** The implementation of the requirement can be determined through basic possible methods: inspection, demonstration, test (instrumented) or analysis (to include validated modelling & simulation).

3.5.1 Mapping Use Cases to Functional Requirements

Table 5 presents the functional requirements with the additional management information for each of the identified functional requirements:

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

- **Requirement ID**: A unique identifier of the functional requirement
- **Requirement Description**: A detailed, high-level description of the requirement
- **Category**: A high-level categorisation based on the functionality required. The elicited MATRYCS functional requirements have been classified in the following categories, where applicable:
    i. Analytics
    ii. Collection
    iii. Compatibility
    iv. Integration
    v. Interaction
    vi. Notification
    vii. Reporting
    viii. Saving
    ix. Visualisation
- **Related Use Cases**: A list of the related use cases from which the requirement was elicited.
- **Related Services**: A list of the related user groups to which the functionality is mapped.
- **Related User Groups**: A list of the related user groups to which the functionality is mapped.

Table 5: MATRYCS Functional Requirements

ID	Requirement Description	Category	Related Use Cases	Related Services	Related User Groups
FR 001	MATRYCS should monitor the results to verify energy and cost savings of executed preventive maintenance measures.	Analytics	UC01_02 + UC04_02	S1.1 + S1.2	Facility Managers, Citizens & Owners
FR 002	MATRYCS should calculate and monitor sustainability aspects for facility performance improvement measures.	Analytics	UC01_02 + UC04_02	S1.1 + S1.2	Facility Managers, Citizens & Owners

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Requirement Description	Category	Related Use Cases	Related Services	Related User Groups
FR 003	MATRYCS should provide a tool for the identification of possible renovation actions to improve energy efficiency in buildings.	Analytics	UC02_01 + UC05_04	S2.1 + S2.2 + S4.2	Citizens & Owners, Designers, ESCOs, Investors
FR 004	MATRYCS should offer a cost assessment tool for the proposed renovation solutions to buildings.	Analytics	UC02_01	S2.2 + S4.2	Citizens & Owners, Designers, Social Housing Providers
FR 005	MATRYCS should provide a regression model engine which generates the optimal mathematical model based on historical data.	Analytics	UC03_01 + UC06_01	S4.1	Citizens & Owners, ESCOs
FR 006	MATRYCS should provide a statistical engine for method selection.	Analytics	UC03_01 + UC06_01	S4.1	Citizens & Owners, ESCOs
FR 007	MATRYCS should provide EPC calculation service.	Analytics	UC03_03 + UC06_03	S1.3	Citizens & Owners, ESCOs, Investors
FR 008	MATRYCS should provide a forecasting tool for energy related data.	Analytics	UC04_01 + UC06_04	S1.1	Utilities, Facility Managers
FR 009	MATRYCS should provide reports on a facility's behaviour.	Analytics	UC04_01	S1.5	Utilities

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Requirement Description	Category	Related Use Cases	Related Services	Related User Groups
FR 010	MATRYCS should provide daily forecast of PV energy production.	Analytics	UC05_01	S1.1	Citizens & Owners, SMEs
FR 011	MATRYCS should provide energy matching optimisation functionality.	Analytics	UC05_03	S1.2	Citizens & Owners
FR 012	MATRYCS should perform aggregations on the datasets.	Analytics	UC07_01	S3.1 + S3.3	Local & Regional Governments and Urban development professionals
FR 013	MATRYCS should perform data comparisons.	Analytics	UC07_01 + UC09_01	S3.1 + S3.3 + S3.2	Local & Regional Governments and Urban development professionals, Designers
FR 014	MATRYCS should provide an investment prioritisation tool, along with the functionality to simulate scenarios and establish the best-case scenario.	Analytics	UC07_02 + UC10_01	S3.1 + S3.4 + S3.3	Policy Makers
FR 015	MATRYCS should provide estimations based on the EPBD.	Analytics	UC08_01 + UC09_02	S1.3	Citizens & Owners, Policy Makers
FR 016	MATRYCS should compare the performance of different buildings.	Analytics	UC08_01	S0.2	Citizens & Owners

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Requirement Description	Category	Related Use Cases	Related Services	Related User Groups
FR 017	MATRYCS should support basic Energy Efficiency Measures calculations.	Analytics	UC09_01	S3.2	Designers, Policy Makers
FR 018	MATRYCS should support verification service of parameters, such as insulation thickness, windows air chamber, "U" values, thermal bridges, etc.	Analytics	UC09_01	S3.2	Designers, , Policy Makers, Owners
FR 019	MATRYCS should support risk assessment service.	Analytics	UC10_01	S3.3	Policy Makers
FR 020	MATRYCS should support comparisons between energy efficiency actions.	Analytics	UC10_01	S3.3	Policy Makers
FR 021	MATRYCS should support a portfolio analysis engine, based on multiple criteria, in order to support policy making.	Analytics	UC11_01	S3.3	Policy Makers
FR 022	MATRYCS should provide the selection of building analytics services from a list of services.	Analytics	UC00_01 + UC00_02	S0.1	Facility Managers, Designers
FR 023	MATRYCS should support statistical analysis of numerical datasets.	Analytics	General		

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Requirement Description	Category	Related Use Cases	Related Services	Related User Groups
FR 024	MATRYCS should support data analytics with both private and public datasets.	Analytics	General		
FR 025	MATRYCS should generate EPCs based estimations at urban level.	Analytics	UC09_02	S1.3	Policy Makers
FR 026	MATRYCS should provide a template with a list of variables (Energy Conservation Measures etc.) to be specified/selected by the user.	Collection	UC01_01 + UC01_02 + UC03_01 + UC03_03 + UC04_01 + UC04_02 + UC05_02 + UC05_03 + UC05_04 + UC06_01 + UC06_03 + UC06_04 + UC07_01 + UC07_02 + UC09_02 + UC10_01 + UC11_01	S1.2 + S1.4 + S4.1 + S1.3 + S1.1 + S1.5 + S2.1 + S3.1 + S3.2 + S3.3	Facility Managers, Citizens & Owners, ESCOs, Local & Regional Governments and Urban development professionals, Policy Makers, Investors

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Requirement Description	Category	Related Use Cases	Related Services	Related User Groups
FR 027	MATRYCS should support the upload of datasets by users.	Collection	UC01_01 + UC01_02 + UC03_01 + UC03_02 + UC03_03 + UC04_01 + UC04_02 + UC05_02 + UC05_03 + UC05_04 + UC06_01 + UC06_02 + UC06_03 + UC06_04 + UC07_02 + UC08_01	S1.2 + S1.4 + S4.1 + S2.1 + S1.3 + S1.1 + S1.5	Facility Managers, Citizens & Owners, ESCOs, Investors, Utilities, Policy Makers
FR 028	MATRYCS should support 2D/3D data collection.	Collection	UC00_01 + UC00_02	S0.1	Designers, Owners, Facility Managers
FR 029	MATRYCS should support interoperability with APIs for data export.	Collection	General		
FR 030	MATRYCS should provide anonymisation mechanism for personal data.	Collection	General		
FR 031	MATRYCS should support data cleaning operations.	Collection	General		
FR 032	MATRYCS should provide dataset aggregation mechanisms for data from different sources.	Collection	General		

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Requirement Description	Category	Related Use Cases	Related Services	Related User Groups
FR 033	MATRYCS should provide privacy restriction mechanisms for uploaded datasets.	Collection	General		
FR 034	MATRYCS should enable the user to insert text input (such as target values, model restrictions etc.).	Collection	UC05_01 + UC05_02 + UC05_03 + UC07_02 + UC10_01 + UC11_01	S1.1 + S3.3	Citizens & Owners, SMEs, Policy Makers
FR 035	MATRYCS should enable the loading of information (such as weather data) from other platforms and APIs	Collection	UC05_01 + UC05_02 + UC08_01	S1.1 + S1.4 + S1.3	Citizens & Owners, SMEs
FR 036	MATRYCS should support data import from .xml files.	Collection	UC09_01	S3.2	Designers, Policy Makers, Citizens & Owners
FR 037	MATRYCS should provide a querying service towards the digital twin.	Compatibility	UC02_01 + UC03_03 + UC06_03 + UC08_01 + UC09_01 + UC09_02	S2.2 + S4.2 + S1.3 + S3.2	Citizens & Owners, Designers, ESCOs, Investors, Designers, Policy Makers
FR 038	MATRYCS should support connection to GUIs	Compatibility	UC03_03 + UC06_03	S1.3	Citizens & Owners, ESCOs, Investors
FR 039	MATRYCS should support connections to web services and APIs.	Compatibility	UC05_01 + UC05_02 + UC08_01	S1.1 + S1.4 + S1.3	Citizens & Owners, SMEs

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Requirement Description	Category	Related Use Cases	Related Services	Related User Groups
FR 040	MATRYCS should combine existing data and user-uploaded data from different sources.	Compatibility	UC07_01 + UC08_01 + UC09_02	S3.1 + S3.3 + S1.3 + S3.2	Local & Regional Governments and Urban development professionals, Citizens & Owners, Policy Makers
FR 041	MATRYCS should support different types of datasets (historical, live, etc.).	Compatibility	UC00_01 + UC00_02	S0.1	Designers
FR 042	MATRYCS should utilise the digital twin.	Compatibility	UC00_01 + UC00_02	S0.1	Designers
FR 043	MATRYCS digital twin functionality should be interoperable with analytics tools.	Compatibility	UC00_01 + UC00_02	S0.1	Designers
FR 044	MATRYCS should perform data sources integration in order to develop a building's digital twin.	Integration	UC00_01 + UC00_02	S0.1	Designers
FR 045	MATRYCS should use an advanced ontology.	Integration	UC00_01 + UC00_02	S0.1	Designers
FR 046	MATRYCS should enable the user to create a new project/activity.	Interaction	UC01_01 + UC01_02 + UC03_01 + UC03_03 + UC04_02 + UC05_02 + UC06_01 + UC06_03	S1.2 + S1.4 + S4.1 + S1.3	Facility Managers, Citizens & Owners, ESCOs, Investors
FR 047	MATRYCS should support building identification operation.	Interaction	UC02_01	S2.2 + S4.2	Citizens & Owners, Designers

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Requirement Description	Category	Related Use Cases	Related Services	Related User Groups
FR 048	MATRYCS should support user interaction with the platform, enabling data editing.	Interaction	UC02_01	S2.2 + S4.2	Citizens & Owners, Designers
FR 049	MATRYCS should support filtering for visualisation tools.	Interaction	UC07_01	S3.1 + S3.3	Local & Regional Governments and Urban development professionals
FR 050	MATRYCS should enable service selection by the user.	Interaction	UC00_01 + UC00_02	S0.1	
FR 051	MATRYCS should enable users to search for datasets with various filters.	Interaction	General		
FR 052	MATRYCS should provide alarm messages when anomalies are detected.	Notification	UC01_01 + UC05_02	S1.4	Facility Managers, Citizens & Owners, ESCOs, Investors
FR 053	MATRYCS should provide error messages in case of false user operation.	Notification	General		
FR 054	MATRYCS should inform users for updates on datasets they use.	Notification	General		

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Requirement Description	Category	Related Use Cases	Related Services	Related User Groups
FR 055	MATRYCS should be able to report the proposed portfolio of energy efficiency improvement actions, along with energy, cost and emissions savings of the proposed actions.	Reporting	UC01_02 + UC04_02	S1.1 + S1.2	Facility Managers, Citizens & Owners
FR 056	MATRYCS should provide data categorisation tools.	Reporting	UC03_03 + UC06_03 + UC09_02	S3.2	Policy Makers, Citizens & Owners, ESCOs, Investors
FR 057	MATRYCS should be able to report analysis results.	Reporting	UC01_01 + UC03_03 + UC05_01 + UC05_02 + UC05_03 + UC05_04 + UC06_03 + UC06_04 + UC08_01 + UC09_01 + UC09_02	S1.4 + S1.3 + S1.1 + S1.2 + S1.5 + S2.1 + S0.2 + S3.2	Facility Managers, Citizens & Owners, ESCOs, Investors, SMEs, Designers, Policy Makers
FR 058	MATRYCS should be able to create a list of preventive maintenance actions.	Reporting	UC01_01	S1.2	Facility Managers
FR 059	MATRYCS should be able to create a reliable implementation action plan.	Reporting	UC01_01	S1.2	Facility Managers
FR 060	MATRYCS should create the corrective measures to properly mitigate risks.	Reporting	UC01_01	S1.4	Facility Managers

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Requirement Description	Category	Related Use Cases	Related Services	Related User Groups
FR 061	MATRYCS should provide reports on energy savings and cost savings.	Reporting	UC03_02 + UC05_04 + UC06_02 + UC10_01	S4.1 + S2.1 + S1.4 + S3.3	Citizens & Owners, ESCOs, Policy Makers, Investors
FR 062	MATRYCS should provide reports on emission savings.	Reporting	UC03_02 + UC06_02	S1.4	Citizens & Owners, ESCOs
FR 063	MATRYCS should report the estimated PV performance.	Reporting	UC05_02	S1.1	Citizens & Owners
FR 064	MATRYCS should provide suggestions for increasing self-consumption and energy matching optimisation.	Reporting	UC05_03	S1.1 + S1.2	Citizens & Owners
FR 065	MATRYCS should provide information about cities' emissions and energy consumption.	Reporting	UC07_01	S3.1 + S3.3	Local & Regional Governments and Urban development professionals
FR 066	MATRYCS should provide specific improvement actions for climate change assessment.	Reporting	UC07_01 + UC07_02	S3.1 + S3.3	Local & Regional Governments and Urban development professionals, Policy Makers
FR 067	MATRYCS should report the number of energy efficiency interventions in each of the proposed actions, along with the proposed year of implementation.	Reporting	UC11_01	S3.3	Policy Makers

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Requirement Description	Category	Related Use Cases	Related Services	Related User Groups
FR 068	MATRYCS should be able to automatically save generated data.	Saving	UC01_01	S1.2	Facility Managers
FR 069	MATRYCS should be able to automatically save parameters generated from Machine Learning algorithms.	Saving	UC01_02 + UC04_02	S1.2	Facility Managers, Citizens & Owners
FR 070	MATRYCS should provide advanced visualisation tools to show the generated results.	Visualisation	UC01_01 + UC01_02 + UC02_01 + UC03_02 + UC03_03 + UC04_02 + UC06_02 + UC06_03 + UC07_01 + UC07_02 + UC09_01 + UC09_02	S1.2 + S1.4 + S2.2 + S4.2 + S1.3 + S3.1 + S3.3 + S3.2 + S0.2	Facility Managers, Citizens & Owners, Designers, ESCOs, Investors, Local & Regional Governments and Urban development professionals, Policy Makers
FR 071	MATRYCS should provide visualisations of the opportunities through a geo-clustering tool, maps or charts.	Visualisation	UC07_01 + UC09_02	S3.1 + S3.3 + S3.2 + S0.2	Local & Regional Governments and Urban development professionals, Policy Makers, Citizens & Owners, Designers
FR 072	MATRYCS should be able to visualise energy performance information.	Visualisation	UC08_01	S0.2	Citizens & Owners
FR 073	MATRYCS should provide classification of the generated solutions.	Visualisation	UC10_01	S3.3	Policy Makers

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Requirement Description	Category	Related Use Cases	Related Services	Related User Groups
FR 074	MATRYCS should provide a graphical representation of optimal energy efficiency portfolios.	Visualisation	UC11_01	S3.3	Policy Makers

3.5.2 Mapping Use Cases to Non-Functional Requirements

Table 6 presents the non-functional requirements with the additional management information for each of the identified non-functional requirements:

- **Requirement ID**: A unique identifier (ID) of the non-functional requirement
- **Requirement Description**: A detailed, high-level description of the requirement
- **Category**: A high-level categorisation based on the functionality required. The elicited MATRYCS non-functional requirements have been classified in the following categories, where applicable:
  i. Modularity
  ii. System Availability
  iii. Efficiency
  iv. Interoperability
  v. Reliability
  vi. Security
  vii. Usability

Table 6 MATRYCS Non-Functional Requirements

ID	Requirement Description	Category
NFR 001	MATRYCS components should be operating independently.	Modularity
NFR 002	MATRYCS should provide the proper mechanisms for system upgrade with minimum downtime.	System Availability
NFR 003	MATRYCS should handle Big Data with efficiency in a high processing speed.	Efficiency
NFR 004	MATRYCS should guarantee data storage in an efficient way providing optimal space allocation.	Efficiency

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Requirement Description	Category
NFR 005	MATRYCS should provide time-efficient analytics methods.	Efficiency
NFR 006	MATRYCS should utilise state-of-the-art algorithms with optimal complexity.	Efficiency
NFR 007	MATRYCS should support (near-) real time data collection of streaming data	Interoperability (towards field components)
NFR 008	MATRYCS should support bi-directional data communication	Interoperability (towards field components)
NFR 009	MATRYCS should guarantee data harmonisation for possibly unstructured data coming from heterogeneous data sources	Interoperability (towards field components)
NFR 010	MATRYCS should allow interoperability with other IT platforms	Interoperability (towards other systems)
NFR 011	MATRYCS should be able to recover if system failure occurs	Reliability
NFR 012	MATRYCS should isolate software errors, maintaining its functionality.	Reliability
NFR 013	MATRYCS should guarantee the security of stored information.	Reliability
NFR 014	MATRYCS should be resilient to simultaneous requests.	Reliability
NFR 015	MATRYCS should implement suitable mechanisms for data persistence to avoid data loss	Reliability
NFR 016	MATRYCS should support different authorisation access level for different users.	Security
NFR 017	MATRYCS should be available in case of Denial of Service (DoS) attacks.	Security
NFR 018	MATRYCS must ensure the integrity of the customer account information.	Security

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Requirement Description	Category
NFR 019	MATRYCS should not return restricted pages to unauthorised users.	Security
NFR 020	MATRYCS should implement mechanisms for avoiding data corruption	Security (associated to the use of blockchain)
NFR 021	MATRYCS should enable the processing of datasets of various formats.	Usability
NFR 022	MATRYCS should support error protection methods for user input fields.	Usability
NFR 023	MATRYCS visualisation tools should be coherent and comprehensible for all targets groups.	Usability
NFR 024	MATRYCS should enable easy user navigation through the platform.	Usability
NFR 025	MATRYCS should offer user support (wizards, etc.) during data uploading.	Usability
NFR 026	MATRYCS should provide online user guides to facilitate user navigation.	Usability
NFR 027	MATRYCS should prevent users from inserting invalid input.	Usability
NFR 028	MATRYCS should provide reports in a standardised and automated way.	Usability
NFR 029	MATRYCS should allow scalability	Usability
NFR 030	MATRYCS should allow extensibility and easy implementation of new functionalities	Usability
NFR 031	MATRYCS should support polyglot software implementation	Usability

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

4 MATRYCS Technical Specification

4.1 Technical Requirements Elicitation

The User Requirements (Functional and Non-Functional) have highlighted the needs of users and the expected functionalities that should be fulfilled from the implementation of the MATRYCS framework. The next step of the technical specification identification involves the transformation of User Requirements to Technical Requirements. The scope of Technical Requirements is to describe the internal implementation, behaviour and standards of MATRYCS framework.

The lifecycle of Technical Requirements begins with the representation of business needs as requirements, continues with the development of specific solutions to these needs and ends with the implementation of the product that satisfies the abovementioned requirements.

As described in the Grant Agreement, MATRYCS is based on short iterative cycles of work, with highly parallel streams of activities by adopting a SCRUM-like software engineering methodology. Therefore, in each iteration the technical requirements shall be revised -and if needed redrafted- in order to ensure that the developed technologies are in line with the user expectations.

In this section, a first iteration of MATRYCS Technical Requirements is presented. In Table 7, there is a description of MATRYCS General Technical Requirements along with related User Requirements from which each requirement has been elicited. This table is the result of an iterative procedure, involving all technical partners of the project, in order to ensure the satisfiability of the requirements and the correct mapping to the MATRYCS modules and tasks.

The component-related MATRYCS Technical Requirements are presented in Sections 4.2, 4.3 and 4.4 for MATRYCS GOVERNANCE, MATRYCS PROCESSING and MATRYCS ANALYTICS, respectively.

Table 7: MATRYCS General Technical Requirements

ID	Technical Requirement Description	Related User Requirements
TR050	The MATRYCS platform must provide system modularity, providing the independent operation of its modules	**NFR001**
TR051	The MATRYCS platform must guarantee system recoverability	**NFR011, NFR012, NFR015, NFR017**
TR052	The MATRYCS platform must provide the possibility for future functionalities, ensuring system scalability and extensibility	**NFR029, NFR030**

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Technical Requirement Description	Related User Requirements
TR053	The MATRYCS platform will follow an incremental approach which facilitates the design and planning off the platform's functionalities through an iterative process	NFR029, NFR030
TR054	Each MATRYCS development iteration is focused on fixing and addressing the issues of the previous iteration but at the same time to provide on extending the list of offerings of the platform with new functionalities	NFR029, NFR030
TR055	MATRYCS Components Integration and formulation of the final platform	NFR029, NFR030
TR056	MATRYCS platform functional, integration and stress testing procedures	NFR029, NFR030
TR057	The MATRYCS platform must support software implementation in many programming languages	NFR031

4.2 MATRYCS GOVERNANCE

In Table 8, the Technical Requirements for MATRYCS GOVERNANCE are presented. Each requirement has a unique ID. A description of the requirement is given, along with the related user requirements. Finally, the related tasks and the module of the requirement to be addressed are provided.

Table 8: MATRYCS GOVERNANCE Technical Requirements

ID	Technical Requirement Description	Related User Requirements	Related Task(s)	Related Module
TR001	The MATRYCS platform must provide data cleansing functionalities.	FR031	Task 3.3	Data pre-processing service

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Technical Requirement Description	Related User Requirements	Related Task(s)	Related Module
TR002	The MATRYCS platform must provide a data anonymisation tool / service, in case that sensitive data are processed.	**FR030**	Task 3.3	Data pre-processing service
TR003	The MATRYCS platform must comply with a common data model.	**FR045**	Task 3.3	Data pre-processing service
TR004	The MATRYCS platform must provide secure data storage functionality.	**FR068, FR069, NFR002, NFR004, NFR013**	Task 3.4	Data Storage & Query Engine
TR005	The MATRYCS platform must ensure that external data being imported in MATRYCS are mapped to the MATRYCS data model.	**NFR009, NFR021**	Task 3.3	Data pre-processing service
TR006	The MATRYCS platform must provide a high-performance distributed query engine.	**FR 037, FR042, NFR003, NFR014**	Task 3.4	Data Storage & Query Engine
TR007	The MATRYCS platform storage system must provide optimal space allocation.	**NFR004**	Task 3.4	Data Storage & Query Engine

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Technical Requirement Description	Related User Requirements	Related Task(s)	Related Module
TR008	The MATRYCS platform must provide a blockchain platform to enable secure data and models sharing for cross-stakeholders.	**NFR020**	Task 3.1	DLT & SMC
TR009	The MATRYCS platform must integrate cross-platform heterogenous data.	**FR028, NFR009**	Task 3.2	Interoperability Module
TR010	The MATRYCS platform must collect real time streaming data.	**NFR007**	Task 3.2	Interoperability Module
TR011	The MATRYCS platform must provide interoperability with other systems and platforms.	**NFR008, NFR010**	Task 3.2	Interoperability Module
TR012	The MATRYCS platform must provide semantic annotation of data.	**NFR009**	Task 3.3	Semantic annotation module
TR013	The MATRYCS platform must offer a reasoning engine.	**FR045**	Task 3.5	Reasoning Engine
TR016	The MATRYCS platform must offer a well-defined API for data export.	**FR029**	Task 3.4	Data Storage & Query Engine

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Technical Requirement Description	Related User Requirements	Related Task(s)	Related Module
TR017	The MATRYCS platform must support importing datasets in different formats (CSV, XLSX, etc.).	**FR027, FR035, FR036, FR041**	Task 3.2	Interoperability Module
TR018	The MATRYCS platform must support importing large files.	**FR027**	Task 3.2	Interoperability Module
TR023	The MATRYCS platform must ensure different authorisation levels for accessing datasets.	**FR033, NFR016, NFR019**	Task 3.6	End-to-End Security Framework
TR025	The MATRYCS platform must provide a secure and controlled registration process for new users.	**NFR016, NFR018**	Task 3.6	End-to-End Security Framework
TR029	The MATRYCS platform must provide data editing by authorised users.	**FR048**	Task 3.6	End-to-End Security Framework
TR038	The MATRYCS platform must forbid unauthorised user access to the platform and the datasets.	**FR033, NFR019**	Task 3.6	End-to-End Security Framework

4.3 MATRYCS PROCESSING

In Table 9, the Technical Requirements for MATRYCS PROCESSING are presented. Each requirement has a unique ID. A description of the requirement is given, along with the related user requirements. Finally,

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

the related tasks and the module of the requirement are provided.

Table 9: MATRYCS PROCESSING Technical Requirements

ID	Technical Requirement Description	Related User Requirements	Related Task(s)	Related Module
TR014	The MATRYCS platform must allow data connection with MATRYCS governance layer.	**FR035, FR039, NFR010**	Task 4.5	Virtual Workbench - Building Data & Services Innovation Hub (Data sets submodule)
TR015	The MATRYCS platform must allow the user to upload datasets and ML/AI models.	**FR027**	Task 4.5	Virtual Workbench - Building Data & Services Innovation Hub (Data sets submodule and function catalogue sub-module)
TR019	The MATRYCS platform must show false usage messages when misuse of the platform occurs.	**FR053, NFR022, NFR027**	Task 4.5	Virtual Workbench - Building Data & Services Innovation Hub
TR020	The MATRYCS platform must provide alarm messages.	**FR052**	Task 4.5	Virtual Workbench - Building Data & Services Innovation Hub
TR021	The MATRYCS platform must allow notification preferences management.	**FR052**	Task 4.5	Virtual Workbench - Building Data & Services Innovation Hub
TR024	The MATRYCS platform must enable verification of the identity of all users using the platform.	**FR033**	Task 4.5	Virtual Workbench - Admin settings
TR028	The MATRYCS platform must provide user navigation and functionalities of the platform in a convenient way.	**FR034, FR046, FR047**	Task 4.5	Virtual Workbench
TR030	The MATRYCS platform must support search functionality over the datasets and the services.	**FR049, FR051, FR056**	Task 4.5	Virtual Workbench

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Technical Requirement Description	Related User Requirements	Related Task(s)	Related Module
TR031	The MATRYCS platform must provide model selection.	**FR005**	Task 4.3	Model Serving Module
TR032	The MATRYCS platform must provide tools and services to apply machine learning algorithms	**NFR006**	Task 4.2	Model Development
TR033	The MATRYCS platform must provide tools and services to apply deep learning algorithms	**NFR006**	Task 4.2	Model Development
TR034	The MATRYCS platform must provide data visualisation tools and functionalities (graphs, maps etc.).	**FR070, FR071, FR072, FR073, FR074, NFR023**	Task 4.4	Visualisation Engine
TR035	The MATRYCS platform must provide reporting tools and services to report the generated numerical results.	**FR057, FR060, FR061, FR062, FR063, FR065**	Task 4.4	Visualisation Engine
TR036	The MATRYCS platform must provide services to propose list of actions.	**FR055, FR058, FR059, FR064, FR066, FR067**	Task 4.4	Visualisation Engine + functionalities from WP5
TR039	The MATRYCS platform's usage must be supported by user guidelines, concerning both navigation and data uploading.	**NFR025, NFR026**	Task 4.5	Virtual Workbench
TR040	The MATRYCS platform must offer easy user navigation through the platform.	**FR038, NFR024**	Task 4.5	Virtual Workbench

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

4.4 MATRYCS ANALYTICS

In Table 10 the Technical Requirements for MATRYCS ANALYTICS are presented. Each requirement has a unique ID. A description of the requirement is given, along with the related user requirements. In this case, there are no related tasks and modules, because MATRYCS ANALYTICS refers to the services that will be developed. And therefore, no such categorisation can be made.

Table 10: MATRYCS ANALYTICS Technical Requirements

ID	Technical Requirement Description	Related User Requirements
TR022	The MATRYCS platform must provide user notifications on dataset updates.	**FR054**
TR026	The MATRYCS platform must provide user-friendly templates for variable specification.	**FR026**
TR027	The MATRYCS platform must provide user-friendly interfaces for text input by the users.	**FR034**
TR037	The MATRYCS platform must support report exporting in various formats (pdf, xlsx etc.)	**FR055, FR058, FR059, FR064, FR066, FR067, NFR028**
TR041	The MATRYCS platform must enable the integration and combined analysis over multiple datasets.	**FR016, FR024, FR032, FR040**
TR042	The MATRYCS platform must provide tools/services to define and execute what-if scenarios on the datasets.	**FR014**
TR043	The MATRYCS platform must allow a user to easily perform aggregations on a dataset.	**FR012, FR032**
TR044	The MATRYCS platform must provide tools and services to apply analytics.	**FR002, FR003, FR004, FR005, FR007, FR009, FR015, FR017, FR018, FR019, FR025**

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

ID	Technical Requirement Description	Related User Requirements
TR045	The MATRYCS platform must provide tools and services that enable users to perform statistical analysis.	**FR006, FR023**
TR046	The MATRYCS platform must provide tools and services which provide forecasts and predictions	**FR008, FR010, NFR005, NFR006**
TR047	The MATRYCS platform must provide tools that apply advanced optimisation models.	**FR011, FR021, NR005**
TR048	The MATRYCS platform must provide tools and services that enable data comparisons.	**FR013, FR016, FR020**
TR049	The MATRYCS platform must provide tools that provide monitoring of the proposed actions.	**FR001, FR002**
TR058	The MATRYCS platform must allow service selection by the user.	**FR022, FR050**
TR059	The MATRYCS platform must provide the development of buildings' digital twin.	**FR044**
TR060	The MATRYCS platform must provide interoperability between digital twin and other analytics services.	**FR037, FR043**

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

5 MATRYCS Security Specification

As part of the development of the MATRYCS framework, a large amount of data from either LSPs or external sources will be used, as indicated in section 2.3 Data Acquisition. Thus, it is necessary to take into account all the aspects that will ensure the data security at all stages of development up to the final product. In addition, the use of datasets by MATRYCS technical partners and also the creation of new technologies and services raises ethical issues as well as intellectual property rights. All the above-mentioned issues are tackled in the following section. More specifically, in section 5.1 the Intellectual Property Rights of the MATRYCS project are analysed, as well as the Ethical Issues associated with the implementation of the project. In section 5.2 the Network and Information Security requirements of the project are presented. Finally, in section 5.3 the GDPR regulatory framework is described.

5.1 IPR and Ethical Issues Requirements

The implementation of the MATRYCS project builds upon several background and open-source components and it is expected to generate multiple new components, modules, software programs and information that can be eligible for intellectual property protection. Moreover, being a project about Big Data, MATRYCS will make use of large amounts of different types of data coming from heterogeneous sources and providers. For this reason, appropriate Intellectual Property Rights (IPR) management strategies are required and the terms of use of the data provided for the project execution need to be clearly expressed. In addition, particular attention must be paid to identifying and addressing the potential ethical issues associated to the collection and processing of the provided data.

This section aims at providing the overview of the obligations and regulatory framework that applies in terms of IPR and ethical issues management, which will serve as a basis for the project partners as well as for the development of specific tasks of the project, such as the implementation of the needed data security solutions in the different MATRYCS layers and the preparation of dissemination activities and exploitation plans. In the following, a list of definitions for the terms commonly used in this context is first provided. Then, an overview of the IPR management strategy (already partially anticipated in the Grant Agreement) is given, followed by details on the key aspects of the IPR management and on the measures that will be taken to mitigate the potential ethics risks associated to the project data.

Table 11: IPR Main definitions

Main definitions	
Term	**Meaning**
Background	Data, know how or information of any kind and form, either tangible or intangible, including intellectual property rights, that were already owned by the project partners before the start of the project and that are necessary for the implementation of the tasks and actions foreseen in the project.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

Main definitions	
Results/foreground	Tangible or intangible goods created as outcome of the project, here including data, software programs, tools, knowledge or information of any kind and form, either tangible or intangible, regardless of whether they are protectable or not.
Intellectual Property Right	Legal rights aimed at protecting the creations of the intellect, such as inventions, the appearance of products, literary, artistic and scientific works and signs, among others.
Access rights	The right to use results or existing knowledge and property rights, according to the conditions set out in the Grant and Consortium Agreement and as further specified in the IPR management strategy.
Exploitation	Use of results in research activities outside the project, in the development, manufacture and marketing of a product or process, in the development and provision of a service or in standardisation activities.
Dissemination	Disclosure of the results by appropriate means (other than protection or exploitation of the results), including presentations, reports, scientific publications in any kind of media, with the objective of sharing them with potential users and peers.

5.1.1 IPR Management

The IPR management must be conducted in accordance with the Model Grant Agreement and the associated guidelines defined in its Chapter 4. In particular, the key aspects to be addressed are:

- Access rights
- Results ownership
- Results protection
- Exploitation
- Dissemination

To this purpose, along with the exploitation strategy performed in WP7, the consortium will create and

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

maintain an IPR Directory throughout the lifetime of the project. The IPR Directory will be a live document where the agreements regarding the partners' ownership and access rights for all the intellectual properties arising within the project will be recorded. This document will include the details related to the foreground created in the project, explicitly defining its status, nature, ownership as well as the associated dissemination and protection measures. In particular, for each intellectual property, the following aspects will be identified:

- Nature and ownership
- Access rights for project partners during the project lifetime
- Rights of use after the end of the project
- Specific requirements associated to the IPR management
- Plans for the exploitation
- Measures for intellectual property protection
- Licensing considerations

5.1.1.1 Access rights

The term "access rights" refers to the rights granted for accessing the results or the already existing knowledge. The access rights determine which parties can use which pieces of foreground/background, for research purposes and/or for exploitation purposes, and on what conditions. In MATRYCS, in general, methodology documents, case studies and developed tools will be available to all the project beneficiaries and shared through a common repository. More in details, at this stage, the following preliminary agreements have been reached:

- **Access to background**: the project partners will give each other access to the background needed for the implementation of the different tasks and for the development of the tools and components needed within the MATRYCS platform. If necessary, non-disclosure agreements may be put in place for some of the background shared during the project.
- **Access to tools and algorithms**: MATRYCS will develop several tools and algorithms as part of the different platform layers; during the project, these tools and algorithms may be available at the discretion and under the terms defined by the proprietary partner. However, technological partners will always share all the information and knowledge that is needed for the coordination of the activities and for the implementation of the connected and/or subordinated tasks. After the end of the project, arrangements can be made, if required, to grant access to the developed tools and algorithms for the exploitation activities of other project partners. In this case, the arrangements will be stipulated following the principle of fair and reasonable conditions outlined in the Grant Agreement, taking into account type of request, value of the tool/algorithm, duration of the access, and terms of the usage. At the same time, it is worth highlighting that the consortium is favourable and willing to adopt open access policies for many of the developments in the project. Many software libraries that are at the basis of the developed tools and algorithms will be therefore released as open source.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

- **Access to pilot data**: data are the core of the MATRYCS project. MATRYCS builds upon the usage of large datasets coming from different pilots and associated to a variety of data sources and scales. Specific data management procedures are required, and have been already identified, to ensure data protection and compliancy with the GDPR regulation. This includes the definition of the roles of data processor and data controller, as it will be further described in Section 5.3.1. Since data management is critical, in particular for those scenarios where personal or sensitive data are present, the access to the raw pilot data will be granted only to the technical partners responsible for the development of the MATRYCS components and services. The possibility to grant access, also for third parties, to the anonymised data and to those datasets that do not present risks in terms of data protection will be evaluated case by case jointly with the data providers.
- **Access to the MATRYCS platform as a whole**: the overall MATRYCS platform will be the main outcome of the project. As already stipulated in the Grant Agreement, the platform will be openly accessible to everybody, also outside the consortium, during the whole lifetime of the project. After the end of the project the intention is to maintain the MATRYCS platform and services to an open-source cloud infrastructure, the access rights to the platform, or to specific components of it, may be reviewed in light of the individual and joint exploitation actions that will be planned to make the MATRYCS solution sustainable in the medium and long term.
- **Access to other results**: during the project lifetime, other results (such as research results, knowledge, etc.) will be shared and made accessible to all the project partners. Moreover, where and when meaningful, ad hoc dissemination activities will be put in place to ensure the timely transfer of knowledge towards a broader audience, in order to facilitate the wide adoption of project results.

The following Table summarises the preliminary agreement in terms of IPR and access rights for the main foreground items expected to be generated in the project.

Table 12 IPR and Access Rights

Initial agreement on IP and use rights	Technology partners	Other partners
MATRYCS GOVERNANCE	IPR	Use rights
MATRYCS PROCESSING	IPR	Use rights
MATRYCS ANALYTICS	IPR	Use rights
Pilot data	Use rights	restricted
MATRYCS as a whole	Open access	
Best practices & applications	Open access	

5.1.1.2 Results ownership

With respect to the ownership of the results, the Grant Agreement specifies that "*the results are owned by the beneficiary that generates them*". Therefore, the ownership of the foreground, namely the

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

intellectual property created during the project, will remain with the project partner that carried out the work to generate that foreground. In some cases, it is possible that more than one beneficiary own a foreground item jointly. This can happen if:

- The foreground item has been generated jointly, as a result of collaboration activities.
- It is not possible to establish the precise contribution of each partner involved in the generation of the foreground item and/or to decouple such contributions.

In case of joint ownership, as recommended in the Grant Agreement, the involved partner will stipulate a "joint ownership agreement" to agree on:

- Division of the ownership (equal or not).
- Conditions for granting licenses.
- Agreements on the result protection, exploitation and dissemination.

The transfer of ownership of a specific foreground item is also possible. In this case, all project partners, who may request access rights, need to be informed in advance and they have the rights to object within a specified period. In case the transfer of ownership is accepted, appropriate arrangements need to be defined to make sure that all the obligations (mainly in terms of access rights, but not only) are passed on to the new owner of the result.

In MATRYCS, the PMB will discuss, identify and assign owners for each of the new results generated in the project. Each foreground item and the associated owner will be then reported in the IPR Directory.

5.1.1.3 Results protection

All the project partners must take appropriate measures to protect the foreground items they own, in particular when they are commercially or industrially exploitable. The most important measures to protect the created intellectual property vary depending on the nature of the result and include:

- **Patents**: they aim at protecting technical inventions, namely solutions to specific problems in the field of technology. The requirement for patentability is that the technical invention is industrially applicable, new and non-obvious. Patents consist of a legal document, issued by a government office, which describe the invention and gives the legal rights to commercially exploit such invention exclusively to its owner (inventor). Patents provide protection to technical inventions for a period up to 20 years.
- **Copyrights**: they aim at protecting authors of original and creative intellectual works, such as books, music, art, as well as computer software and architectures, against those who copy their work, namely against those who take, use, reproduce, distribute or broadcast the intellectual work in the form created by the author. The length of the copyright protection may vary depending on the type of intellectual work, but it is in most of the cases 50 or 70 years, after which the intellectual work becomes of public domain.
- **Trademarks**: they are signs, symbols, logos or similar items that individualize the goods of an enterprise, distinguishing them from those of the competitors. Trademarks are protected by law and, to be eligible for protection, they need to be distinctive. Trademarks protection can be protracted over time by renewing it periodically (e.g. every 10 years).

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

- **Industrial designs**: they refer to the protection of the visual (i.e. formal, ornamental or aesthetic) appearance of industrial handcraft goods. The requirement for protection in this case is the novel, distinctive and individual character of the to-be protected product. Industrial designs protection can be obtained for a period of time up to 25 years.
- **Trade secrets**: they refer to specific, private and confidential information that is important to a business and for which the owner applies specific measures to keep it secret because it gives a competitive advantage in the marketplace. To be qualified as such, a trade secret should be commercially valuable and known only by a limited number of persons. Confidentiality agreements can be put in place to obtain the legal protection for trade secrets.

Given the nature of the MATRYCS project, the protection measures more relevant for most of the foreground items foreseen in the project is the copyright. From this perspective, the main requirements and implications associated to the copyright protection and licenses handling were already identified before the start of the project, as also reported in Grant Agreement. The consortium plans to adopt the following measures and to consider the following requirements for the management and protection of the created foreground:

i. The MATRYCS solution will be copyright-protected and, to this purpose, a licensing scheme that does not violate the terms and conditions associated to each of the components comprising it needs to be identified.
ii. The partner generating the foreground holds the ownership of it and, as such, it may identify the licensing options more appropriate to defend its own legitimate interests. At the same time, however, also the possible implications and the legitimate interests of the other project partners should be considered for the final choice of the licensing option.
iii. Eventually, the consortium needs to examine the software licenses associated to each algorithm, tool or component embedded into the platform and decide accordingly under which license, or combination of licenses, the MATRYCS framework will be released.
iv. The consortium embraces the vision that open-source licenses providing unrestricted access to software components are extremely important to foster scientific progress, facilitate dissemination and collaboration with third parties, and for showcasing the project outcomes. The consortium commits to apply open-source policies whenever possible, but, since some of the components may be derived from legacy / proprietary components, these will not be released as open source.
v. For the components that can be delivered as open source, an analysis will be carried out to identify the more convenient open source licensing option, considering the specific clauses behind the usage of each license and the possible limitations that could consequently arise.
vi. For the components that cannot be delivered as open source, they will be copyright-protected but still accessible to all the partners for the generation of their results, as also mentioned in the access rights section.

5.1.1.4 Results exploitation

As required in the Grant Agreement, the project partners must take measures to exploit the generated foreground after the end of the project. Exploitation activities may include the use of the acquired

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

knowledge in new projects or research activities, further developing and eventually marketing goods or products derived from the project work and creating and offering services associated to the project developments.

The exploitation strategies will clearly vary depending on the nature of project partner (academic, industrial, etc.). In general, all the project partners will create individual exploitation plans tailored to their focus and business. In addition, joint exploitation opportunities will be investigated. At this stage of the project, the major assets of MATRYCS that have been already identified as potential exploitable items are:

- The tools, algorithms and modules behind each layer of the MATRYCS platform
- The MATRYCS platform as a whole

The first point comprises a number of different components that are likely to have an individual partner ownership. This thus provides clear opportunities for individual exploitation. The second point refers to the integrated MATRYCS platform, which is an item whose business value has to be evaluated in the framework of a joint exploitation strategy. The detailed analysis of the business modelling and of the exploitation definition is part of a dedicated WP of the project. More details will thus result from the activities planned in the WP tasks and will be provided in the associated Deliverables under WP7 Exploitation and Business Cases development. At the moment of writing of this Deliverable, a first document containing the preliminary business and exploitation ideas (D7.1 Business and Exploitation Plan) has been already issued and is available internally to the consortium.

5.1.1.5 Results dissemination

The project partners must disseminate the project results they generate and make them freely accessible to the audience as soon as possible, unless this goes against the legitimate interests of the result owner. The implementation of an effective dissemination campaign involves different steps, among which:

- Identification of the target groups and engagement of the stakeholders
- Definition of the overall outreach strategy
- Planning of tailored dissemination campaigns
- Identification and use of the most effective communication channels
- Publication of promotional and dissemination material

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

Among the dissemination activities, a key role is played by the scientific publications aimed at disclosing the research results obtained as an outcome of the project. These results must be openly available in order to foster a fast transfer of knowledge also externally to the project (while considering all the measures to be implemented for the protection of exploitable results). MATRYCS will disseminate the MATRYCS research, concepts and provide guided research directions and stimuli via selected events, magazines, workshops and conferences. Project partners will guarantee that scientific publications are made available either via gold access solution (publication on open access journals) or, when this is not possible, via green access solutions (i.e. self-archiving in the institutional repository of the employer). Scientific publications have also to guarantee that do not affect, in any way, the legitimate interests of any of the project beneficiaries. To comply with this requirement, all publications will be first shared internally to the consortium, so that the other partners are given the possibility to raise objections (within a predefined period of time) in case they see a violation of their interests. In addition, open access to the research data should be also provided. In MATRYCS, this possibility is evaluated taking into account the data protection constraints and restrictions that may apply for each pilot, as reported in the section on the access rights.

The definition, planning and implementation of the MATRYCS dissemination activities is analysed more in depth under the WP 8 Communication, Dissemination and Awareness Creation. Accordingly, more details about the dissemination strategies will be provided in the associated Deliverables of this WP.

5.1.2 Ethical issues

MATRYCS will make use of large amounts of different types of data coming from heterogeneous sources and providers. As such, the data processing may entail high risks of ethics issues associated to the following categories:

i. **Data subjects**: this risk is associated to the possible presence of personal data, in specific datasets, related to people who did not give an explicit consent regarding the participation to the project.

ii. **Data processing scale**: it is associated to the possible presence of personal data as well as data coming from sensors used to monitor the access to public areas; moreover, in MATRYCS, the planned activities involve the processing of large amounts of heterogeneous data coming from different data providers, for which high ethic risks arise as well.

iii. **Data collection and processing**: it is associated to the collection of data coming from privacy-invasive sensors; in MATRYCS, the processing performed behind many services will involve use of AI and automated decision-making, which may hence lead to issues if personal or sensitive data are present.

At this stage of the project, a detailed review of the characteristics of the datasets made available by each one of the LSPs is still in progress. Regardless of the outcome of this analysis, the MATRYCS solution must be compliant with scenarios where personal and sensitive data are also present. As a consequence, a major requirement for the platform is to implement all the anonymisation and pseudo-anonymisation measures necessary to remove any possible concern related to ethics. The data management strategy and the associated solutions planned to deal with this issue are presented in Section 5.3 where the full

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

overview of the requirements and measures associated to data protection and compliancy with the GDPR regulation is provided.

5.2 Network and Information Security Requirements

5.2.1 NIS Directive

The present project would like to follow the NIS directive in preparing the requirements. This approach has been adopted in order to take into strong consideration the Network and Information Security (NIS) Directive for all future development and technology transfer after the project completion. The idea under this approach is to have a common understanding of the NIS directive and to spread its adoption to the technical partners during the full project life, specifically during the design and development steps. In this way, the future project results obtained after the end of the funding phase can be easily compliant to the NIS directives and the guidelines or laws, according to the different countries regulatory guidelines. With this approach, all future products and services carried out by the different project partners under the production and commercialisation steps can be compliant with the network and security rules in a better and easier way, even if the NIS directive cannot be directly applied to the internal mock-ups and piloting intermediate results under a R&D project.

To this scope, in the following section a detailed analysis of NIS is presented with a short description also of the main reason to design and launch this directive by the European Commission, including the short history of the directive itself and the relevant status of the directive in the European ecosystem.

5.2.2 Short history of NIS Directive

On 29 January 2020, the European Commission's new work programme was published. Under the second priority - 'A Europe fit for the digital age', the Commission announced its intention to launch a review of the Directive on security of network and information systems (NIS Directive), in order to 'further strengthen overall cybersecurity in the Union'. According to the adjusted work programme, the review should be adopted in the last quarter of 2020.

The current Directive on security of network and information systems entered into force in August 2016. Member States had to transpose it into their national laws by 9 May 2018. The directive lays down requirements regarding national cybersecurity capabilities of Member States; rules for their cross-border cooperation; and requirements regarding national supervision of operators of essential services and key digital service providers.

The Commission launched on 7 July 2020 a public consultation on the revision of the NIS Directive that aims to collect views on its implementation and on the impact of potential future changes. The consultation closed on 2 October 2020.

On 16 December 2020, the European Commission and the High Representative of the Union for Foreign Affairs and Security Policy presented a new EU Cybersecurity Strategy that aims to bolster Europe's collective resilience against cyber threats and ensure that all citizens and businesses can fully benefit from trustworthy and reliable services and digital tools. Accordingly, The Commission made two new proposals: a Directive on measures for high common level of cybersecurity across the Union (revised NIS Directive or 'NIS 2'), and a new Directive on the resilience of critical entities.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

The NIS Directive has increased the EU national cybersecurity capabilities, requiring Member States to elaborate a National Cybersecurity strategy, to establish Computer Security Incident Response Teams (CSIRTs) and to appoint NIS national competent authorities, improving the cyber resilience of public and private entities in specific sectors and across digital services. However, its implementation proved difficult, resulting in fragmentation at different levels across the internal market. In order to respond to the growing threats due to digitalisation and increase in cyberattacks, the proposed revised NIS Directive NIS 2 repeals the existing NIS Directive. The new proposal broadens its scope, aiming to strengthen the security requirements imposed, addressing security of supply chains, streamlining reporting obligations, introducing more stringent supervisory measures and stricter enforcement requirements including harmonised sanctions regimes across Member States. It also includes proposals for information sharing and cooperation on cyber crisis management at national and EU level.

5.2.3 NIS Directive main objectives and scopes

The Directive has three main objectives:

- Improving national cybersecurity capabilities;
- Building cooperation at EU level; and
- Promoting a culture of risk management and incident reporting among key economic actors, notably operators providing essential services (OES) for the maintenance of economic and societal activities and Digital Service Providers (DSPs).

The NIS Directive is a cornerstone of the EU's response to the growing cyber threats and challenges which are accompanying the digitalisation of our economic and societal life, and its implementation is therefore an essential part of the cybersecurity package presented on 13 September 2017. The effectiveness of the EU's response is inhibited as long as the NIS Directive is not fully transposed in all EU Member States. This was also recognised as a critical point in the Commission's 2016 Communication on Strengthening Europe's Cyber Resilience System [18].

The NIS Directive is the first piece of EU-wide legislation on cybersecurity. It provides legal measures to boost the overall level of cybersecurity in the EU. The Directive on security of network and information systems [19] (the NIS Directive) provides legal measures to boost the overall level of cybersecurity in the EU by ensuring:

Member States' preparedness, by requiring them to be appropriately equipped. For example, with a Computer Security Incident Response Team (CSIRT) and a competent national NIS authority, cooperation among all the Member States, by setting up a Cooperation Group to support and facilitate strategic cooperation and the exchange of information among Member States. a culture of security across sectors that are vital for our economy and society and moreover rely heavily on ICTs, such as energy, transport, water, banking, financial market infrastructures, healthcare and digital infrastructure.

Businesses identified by the Member States as operators of essential services in the above sectors will have to take appropriate security measures and to notify relevant national authorities of serious incidents. Key digital service providers, such as search engines, cloud computing services and online marketplaces, will have to comply with the security and notification requirements under the new Directive.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

5.2.4 NISD Guidelines

In the present paragraph a synthesis of the NIS guidelines will be provided to be guidelines to the design of the services and, overall to be baselines for the future products and services implementation after the end of the project in the production phase. These guidelines have been summarised and based on the report provided for DSP and NIS and by ENISA (Guidelines on assessing DSP and OES compliance to the NISD security requirements). The ENISA is the European Union Agency for Network and Information Security, centre of network and information security expertise for the EU, its member states, the private sector and EU citizens. ENISA works with all these groups to develop advice and recommendations on good practice in information security.

First of all, it should be noted that NIS Directive (see Articles 14, 15 and 16), are dedicated to introducing appropriate security measures for operators of essential services (OES) as well as for the digital service providers (DSP). The definition of these operators and the inclusion of them in a specific list is under the member state control and monitoring.

This approach has been chosen in order to achieve a baseline, common level of information security within the European Union (EU) network and information systems. Information security (IS) audits and self–assessment/ management exercises are the two major enablers to achieve this objective.

In order to completion and for a better understanding these main three important NISD articles are in the following reported [20]:

Article 14: *"Member States shall ensure that operators of essential services take appropriate and proportionate technical and organisational measures to manage the risks posed to the security of network and information systems which they use in their operations. Having regard to the state of the art, those measures shall ensure a level of security of network and information systems appropriate to the risk posed."*

Article 15: *"Member States shall ensure that the competent authorities have the powers and means to require operators of essential services to provide (b) evidence of the effective implementation of security policies, such as the results of a security audit carried out by the competent authority or a qualified auditor and, in the latter case, to make the results thereof, including the underlying evidence, available to the competent authority."*

Article 16: *"Member States shall ensure that digital service providers identify and take appropriate and proportionate technical and organisational measures to manage the risks posed to the security of network and information systems which they use in the context of offering services referred to in Annex III within the Union. Having regard to the state of the art, those measures shall ensure a level of security of network and information systems appropriate to the risk posed and shall take into account the following elements: a) the security of systems and facilities, b) incident handling, c) business continuity management, d) monitoring, auditing and testing, and e) compliance with international standards."*

5.2.5 NIS specific requirements

After the general introduction, hereinafter a detailed description of the NISD main requirements is presented organised in different tables:

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

Table 13: MATRYCS NISD Requirements

1.	**Information System Security Risk Analysis**
1.1.	The key personnel should be aware of the main information security risks and the relevant mitigations
1.2.	there should be in place a mechanism for ensuring that all security personnel use the risk management methodology and tools
1.3.	the risk management methodology and/or tools, should be periodically reviewed, taking into account changes and past incidents
2.	Information System Security Policy
2.1.	It should be put in place an information security policy (ISSP) and an information security management system (ISMS)
2.2.	Some certifications for specific security risk management standards could be put in place
2.3.	Some information security processes should be reviewed at regular intervals, taking also into account violations, exceptions and incidents which affected other essential operators/ DSP
3.	Information System Security Accreditation
3.1.	the systems supporting essential services should be regularly subjected to security scans and they should be integrated within the risk management framework of the organisation
3.2.	there should be policy/procedures in place for the performance of security assessments and security testing
3.3.	The effectiveness of policy/procedures for security testing should be carefully evaluated
4.	Information System Security Indicators
4.1.	The KPIs implemented in systems supporting essential services should be able to be assessed versus their effectiveness at all times
4.2.	Policy/procedures should be put in place for the implementation of security indicators for testing the systems supporting essential services
4.3.	The aforementioned policy/procedures should be periodically reviewed and updated
5.	Information System Security Audit
5.1.	An updated policy and/ or procedure for performing information system security assessments should be put in place, including audits of systems and assets supporting essential services
6.	Human Resource Security
6.1.	The professional references of key personnel (system administrators, security officers, guards, et cetera) should be validated
6.2.	Training material on security issues should be provided to key personnel
6.3.	Key personnel should be formally appointed in necessary security roles
6.4.	Policies/procedures for the Human Resource security should be regularly reviewed and updated, taking into account any possible changes
7.	Asset Management
7.1.	Detailed lists of critical assets and configurations of systems supporting essential services should be regularly maintained
7.2.	Policy/procedures should be put in place for asset management configuration control
7.3.	The asset management policy should be regularly updated, on the basis of changes and past incidents
8.	Systems Configuration
8.1.	Networks and systems supporting essential services should be configured with information security in mind
8.2.	The effectiveness of the security configurations to protect the integrity of systems should be regularly evaluated and reviewed

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

9. System Segregation
9.1. The information systems should be properly segregated in order minimize the potential consequences when risks occur
10. Traffic Filtering
10.1. A monitoring mechanism of the systems supporting essential services should be put in place
10.2. A traffic monitoring policy of the systems supporting essential services should be defined and put in place
10.3. Specific tools should be defined for supporting the traffic monitoring of the systems supporting essential services
11. Cryptography
11.1. Cryptographic mechanisms should be put in place to protect the confidentiality and integrity of information stored in or out of the company boundaries (digital facilities)
11.2. Implemented cryptographic mechanisms such as digital signatures and hashes to detect unauthorised changes to critical data at rest should be considered and define
12. Administration Accounts
12.1. The operator should set up specific administration accounts, to be used only for administrators that are carrying out specific operations (e.g. installation, configuration, management, maintenance, etc.) on the systems supporting essential services
12.2. The administrator accounts should be solely used to connect to administration information systems?
13. Administration Information Systems
13.1. Administration information systems should be solely used for administration purposes and not mixed up with other operations
13.2. The aforementioned resources should be managed and configured by an authorised operator
14. Authentication and Identification
14.1. Some access control mechanisms should be defined and put in place, for network and information systems, in order to allow only authorised use
14.2. All unused or no longer needed accounts should be deactivated
14.3. A mechanism should be defined and put in place for monitoring access to network and information systems and for approving exceptions and registering access violations
15. Access Rights
15.1. Access rights granted in a structured and monitored manner should be defined
15.2. The operator should define access rights to the multiple functionalities of the resource
16. IT Security Maintenance Procedure
16.1. Some procedure should be established for security maintenance in accordance with the security policy
16.2. The conditions for enabling the minimum security level for systems supporting essential services resources should be defined
16.3. Software and hardware resources should be regularly maintained and updated
17. Industrial Control Systems
17.1. Considering that the proper operation of many essential services depends on functioning and secure industrial control systems (ICS), all operators, if applicable, should be taken into account the particular security requirements for ICS

5.2.6 NIS selected standards

In addition to the analysis and the guidelines already presented a short summary of the main well-known selected international self-risk assessment/management standards and frameworks is presented in this

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

paragraph. This analysis can be another reference for the MATRYCS team in order to better understand all possible guidelines for the network and security requirements to follow in the project results implementation. The specific analysis describing their main features of these standards has been extracted from the ENISA document and is presented hereinafter. The standard analysed are the following:

- **ISO/IEC 27001** framework for an ISMS;
- **NIST** Special Publication 800-30 Rev. 1, Risk Management Guide for Information Technology Systems;
- **CRAMM** risk management methodology;
- **OCTAVE**, suite of tools, techniques and methods;
- **FAIR**, international standards quantitative mode;
- **IRAM2**, end-to-end approach for performing business-focused information risk assessments;

A short description for all these methodologies is presented in the following paragraphs:

5.2.6.1 ISO/IEC 27001 - ISMS

ISO/IEC 27001 is the international standard for information security management systems (**ISMS**). The ISO/IEC 27001 Standard provides a methodology which can assist OES and DSP to achieve all of their regulatory compliance objectives concerning the NIS Directive by implementing specific controls. Controls recommended by ISO/IEC 27001 are not only technological solutions but also cover people and organisational processes. There are 114 controls in Annex A covering the breadth of information security management, including areas such as physical access control, security staff awareness programmes, procedures for monitoring threats and incident management processes.

The risk assessment process established by ISO/IEC 27001 follows the below procedure:

- establish and maintain certain information security risk criteria;
- ensure that repeated risk assessments "produce consistent, valid and comparable results;
- identify risks associated with the loss of confidentiality, integrity and availability for information within the scope of the information security management system;
- identify the owners of those risks; and
- analyse and evaluate information security risks according to certain criteria.

An ISMS is based on the outcomes of a risk assessment based on the ISO/IEC 27001. OES and DSP will need to produce a set of controls so as to minimize the identified risks resulting from the aforementioned procedure.

5.2.6.2 NIST Special Publication 800-30

NIST Special Publication 800-30 is a foundation pillar for developing an effective and adequate risk management program. NIST 800-30 provides both the definitions and the practical guidance required for assessing and mitigating risks identified within IT systems. Additionally, it provides information on the selection of practical and profitable security controls that can be utilised to mitigate risk for the better protection of vital information and the IT systems that process this

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

information. It is composed by well-defined and sequentially steps in order to achieve the aforementioned goals, as depicted below:

- System characterisation followed by thread and vulnerability identification; control analysis and likelihood determination;
- Impact analysis and risk determination; and
- Control recommendations and documentation of the results.

5.2.6.3 CRAMM

CRAMM (CCTA Risk Analysis and Management Method) was developed in 1987 by a British government organisation, the Central Communication and Telecommunication Agency (CCTA), now renamed into Cabinet Office. CRAMM can be used for all kinds of organisations, but it is especially intended for large organisations, like government bodies and industry50. It is in use by NATO and corporations working actively on information security. CRAMM helps in justification of security investments by demonstrating need for action at management level, based on quantifiable results and countermeasures from organisation.

CRAMM attempts a qualitative approach that focuses on assets. It provides 10 specific and predefined asset tables which classify the assets in categories. Those tables support identification and valuation of assets51. Therefore, each asset can be classified into a specific category, each with a predefined list of known vulnerabilities and threats that can exploit them. After the completion of identification and valuation of the assets, the provided dedicated tool automatically suggests a set of all possible countermeasures. However, the usefulness of the method is largely dependent on the tool which implements it.

5.2.6.4 OCTAVE

OCTAVE (Operationally Critical Threat, Asset, and Vulnerability Evaluation) was developed by the Computer Emergency Response Team within the Software Engineering Institute. The goal of the OCTAVE suite of tools, techniques and methods is to allow "risk-based information security strategic assessment and planning"49. OCTAVE gives the opportunity to small teams across business units and IT work together to address the security needs of the organisation and face the security challenges. It moves an organisation towards an operational risk-based view of security and addresses technology in a business context.

The methodology is divided in three explicit methods. The primary OCTAVE method forms the basis for the OCTAVE foundation of knowledge. OCTAVE-S is intended for small and medium sized organisations. The main difference with the basic method is that the necessary knowledge is assumed to be known in advance by the analysis group, so the first step of collecting knowledge is omitted. Lastly, OCTAVE-Allegro offers a faster but more limited approach that focuses on information assets. This approach covers only four simplified steps: development of risk measurement criteria, creation of profiles for each critical information asset, identification of threats to these assets and finally, analysis of resulting risks in order to develop mitigation approaches.

5.2.6.5 FAIR

FAIR (Factor Analysis of Information Risk) is an international standard quantitative model for information

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

security and operational risk and provides (a) a model for understanding, analysing and quantifying information risk in financial terms; and (b) a foundation for developing a robust approach to information risk management.

The FAIR framework defines the necessary building blocks for implementing effective risk management programs. FAIR is an ontology of the factors that contribute to risk and how they affect each other. It is primarily concerned with establishing accurate probabilities for the frequency and magnitude of data loss events.

5.2.6.6 IRAM2

IRAM2 (Information Risk Assessment Methodology 2) is a complete end-to-end approach for performing business-focused information risk assessments.

- Simple, practical, yet rigorous risk assessment approach; focus on the business perspective;
- Extended coverage of risks; and
- Engagement with key stakeholders.

IRAM2 is supported by four IRAM2 Assistants, each accompanied by a practitioner guide, that help automate one or more phases of the methodology.

These methodologies are the most notable in the field of information security for risk assessment and management. A summary table is presented in the following, comparing the standard with a set of key criteria, specifically:

i. **Scope/ domain:** defines the scope and the domain of applicability of the methodology;
ii. **Focus (RA/ RM):** defines the focus of the methodology, i.e. risk assessment, risk management or both;
iii. **Flexibility:** refers to the flexibility of the methodology;
iv. **Complexity:** refers to the complexity of the methodology;
v. **Approach:** refers to the approach of the methodology;
vi. **Tool support:** defines whether there is a tool which implements the methodology;
vii. **Year released/ last update:** refers to the release year and the last update of the methodology; and
viii. **Target:** refers to the sector and/or the types of entities that are in the scope of the methodology**.**

Table 14: Criteria vs. Methodologies

S/N	CRITERIA	ISO 27001	OCTAVE	CRAMM	FAIR	IRAM2	NIST 800-30
1	Scope/ Domain	SME / LE	LE	SME / LE	SME / LE	LE	SME / LE
2	Focus	RA / RM	RA / RM	RA	RA	RA / RM	RM

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

S/N	CRITERIA	ISO 27001	OCTAVE	CRAMM	FAIR	IRAM2	NIST 800-30
3	Flexibility	Relatively flexible	Flexible	No Flexible	Relatively flexible	Flexible	Relatively flexible
4	Complexity	Medium	Low	High	Low	Low	Low
5	Approach	Assets and control based	Risk based information security strategy	Qualitative, asset-centric approach	Quantitative approach by filling a questionnaire table	Assessment of risk from a business perspective	Risk based It related risk management
6	Tool support	No	yes	yes	yes	yes	N/A
7	Year released /Last update	2005/ 2013	1999/ 2005	1985/ 2011	2001/ 2009	2014/ 2014	2000/ 2012
8	Target	All NISD sector	All NISD sector	All NISD sector	All NISD sector	All NISD sector	All NISD sector

5.3 GDPR Requirements

In recent years, the number of data used for processing for research purposes has been increasing. At the same time, the ability to collect, disseminate, process, and store personal data considered "sensitive data" has been improved at a very low cost through innovative tools and techniques. Nevertheless, improving data processing techniques and increasing the volume of data has the effect of reducing data control, which can lead to abuse and risk to the individual's privacy. The use of personal data without permission is a criminally prosecuted and irreversible act, which violates fundamental human rights.

In this context, the main object of the MATRYCS project is the collection, processing, storage, and possible exploitation of a large amount of personal data from numerous sources, through various technologies and mechanisms to be developed in the individual tasks of the project. Therefore, in order to avoid the misuse of personal data, all actions carried out by technical partners must be compatible with the European legal framework for the protection of individuals' personal data, and in particular it is important that any technological progress in the context of the project contains security and privacy requirements.

This section aims to provide an overview of the existing European regulatory system in order to give the project's technical partners a basis on which to rely in the development of the tools and services they

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

will develop. Firstly, the basic concepts and definitions related to the protection of sensitive data are given. Then, the general principles of European legislation regarding the management of personal data are analysed. Finally, the methodology to be followed in the next stages of the MATRYCS project is described and the possible risks that exist are explored.

5.3.1 Definitions

In this paragraph, the basic definitions of GDPR legislations are given, according to EU Law on personal data and General Data Protection Regulation (GDPR) protocols (EU, 2016/679 Regulation) [21].

Personal data are defined as information related to an identified natural person, and more specifically, a person that can be distinguished from all other members of the society and can be recognised as an individual. The identification of a person can be achieved by combining some personal information such as a name, an identification number, location data, an online identifier etc. It can also be achieved by collecting information about one or more factors related to the physical, physiological, genetic, economic or social identity of that natural person. The person who is involved in the data processing procedure is called **data subject**.

Under EU law, and more specifically the e-Privacy Directive, any data used by an electronic communications service revealing the geographic position of a publicly available electronic communications service's user is called **location data**. Collecting and processing this kind of data requires that technical partners comply with the respective EU legislation. It is not known yet if in the context of the MATRYCS project such data will be exploited; however, it is very likely that location data in a direct or an indirect way will be needed.

There are, also, some special categories of personal data which involve a higher risk compared to normal personal data. The risk stems from the fact that if these data are mishandled, there might be some irreversible and long-term consequences for the data subject, such as the violation of the fundamental right to privacy. These special categories of data are commonly referred as **sensitive data** and it is essential that enhanced protection is imposed against them. Examples of these categories are the following: data which include ethnicity, origin and religious beliefs, personal data that reveal health and/or sexual orientation, genetic and biometric data etc.

The protection of these categories of data is of high importance, therefore the processing of these data is not allowed except from very specific circumstances. In this context, in case that it is needed to analyse, distribute or store such data during the implementation of the MATRYCS toolbox, a specialised research must be conducted according to the type of data in order to secure compliance to the EU and local legislation.

The MATRYCS framework aspires to apply a series of anonymisation techniques in order to generate anonymised data from the personal data that will be collected from the pilots. **Anonymised data** refer to information, which does not relate to -or imply- an identified or identifiable natural person. The procedure of anonymisation can be thought of a series of steps which result in the alteration of a dataset of personal data, in a way that all elements which help the identification of a person are transformed. As a result, the data subject is no longer identifiable by the data. According to EU legislation for GDPR, anonymised data are not subject to the principles of data protection, they are not considered as personal or sensitive data anymore and consequently their processing for statistical or research purpose is not restricted.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

On the other hand, the term **pseudoanonymised data** refers to personal data which have their identifiable attributes transformed by symbols, thus concealing the sensitive data in a way that some additional information is required in order to identify the data subject. The additional information is saved separately, making the identification of a person a quite difficult procedure. Pseudoanonymisation enables the collection of additional personal data of the same data subject, without revealing his/her identity. However, it is significant to note that pseudoanonymised data do not guarantee the anonymisation and untraceability of the subject's identity. Therefore, they are subject to the EU data protection law.

MATRYCS aims to apply anonymisation techniques to every dataset including personal or sensitive data. In case anonymisation is not possible for some datasets, pseudoanonymisation techniques will be applied in order to assure that personal data are protected and the risks for MATRYCS data subjects are diminished.

The term **data processing** includes any action or set of actions performed on personal or sensitive data or datasets of such data. Examples of such actions are data collection, data recording and data storage, data alteration (anonymisation or pseudoanonymisation), data –or results based upon data- dissemination, structuring, combination, or any other use. The term data processing includes both any automated operation which is performed on online, streaming personal data and any non-automated operation which is performed specifically on a subset of the dataset. The scope of MATRYCS is to encapsulate an integrated approach to process Big Data. Therefore, there is a wide set of activities included in the data processing definition such as data collection and storage, data enrichment and querying, data anonymisation etc. It is important to note that the term data processing may refer to non-personal data. However, in this section the use of personal data is highlighted.

There are two categories of users that handle personal data: data controllers and data processors. On the one hand, the role of the **data controller** is the determination of the purpose for which and how personal data is processed, deciding "why" personal must be processed and "how" personal data will be processed. On the other hand, the **data processor** performs the necessary data processing on behalf of the data controller.

In MATRYCS, the existence of multiple LSPs which are involved in different use cases leads to the solution of appointing data controllers, case-by-case. In this context, LSPs will be the **MATRYCS Data Controllers**, thus being responsible of having the permission to use any kind of data for MATRYCS purposes. Therefore, if they have the permission, then they could share data with the technical partners of MATRYCS (EURAC, COMSENSUS, ENG, COOPERNICO, CARTIF, NTUA, HOLISTIC), who have the role of **MATRYCS Data Processors.** The technical partner who is involved in the development, structuring and distribution of each use case is the sole data processor.

Finally, a basic term that enables the processing of personal data is consent. According to EU law, **consent** of the data subject includes "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". In the context of MATRYCS, this principle means that the data controller should inform the data subjects about the operations which will be conducted on the collected data, enabling the data subject to give or not give consent. The consent does not enable the data processor to perform any other operation on data except the agreed ones. If any other need for data operations emerges, then the data controllers should inform the subject on the alterations.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

5.3.2 Personal Data Processing Principles

The main principles which are related to personal data processing are listed in Article 5 of the GDPR [22] which is presented in the following section:

Personal data shall be:

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

Firstly, the principle of **lawfulness, fairness, and transparency**, as described in paragraph 6 of the GDPR law [23], includes the need for consent, the performance of a contract with the data subject, the compliance with any legal obligations, the protection of vital interests as well as the public interest and the overriding of the controller's interest.

The principle of **limited purpose** involves setting limitations on the extent that controllers can use the collected data, more specifically imposing the necessary restrictions if the scope of usage exceeds the agreed purpose. The purpose limitation also affects the transparency and legal certainty of the process.

The principle of **data minimisation** states that the data controllers should only collect personal data which are necessary and relevant in order to accomplish a specific scope and to generate a specified result. Additionally, data controllers should not keep the data for longer than it is necessary for the fulfilment of the original purpose.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

Data accuracy is the principle which ensures that data are kept up to date. Data should be tactically checked in order to ensure that they are not outdated or distorted. Especially in MATRYCS, the streams of online data necessitate the need for data accuracy.

The fifth principle is **storage limitation**. This principle aims to control that the duration of storage does not exceed the period for which the data is necessary for fulfilling the necessary purpose. In MATRYCS, the storage process of the collected data is divided into two parts. Firstly, the test datasets which will be needed for the construction will be stored until the final implementation of the models. Secondly, the data that will be imported after the development of the MATRYCS Toolbox are subject to the user's needs and may be stored for as long as he/she wishes.

Finally, the last set of principles includes **data security, integrity and confidentiality**. In the context of MATRYCS, a series of appropriate technical and organisational measures will be developed in order to assure data and platform security, taking into the high risks of a novel platform, with high cost of implementation and maintenance, as well as the importance of the collected datasets from 11 large scale pilots. Additionally, in case that a data breach occurs, the data processors should immediately notify the personal data breach to the supervisory authority competent. If the data breach is likely to involve high risk of illegal interception of sensitive data, then the data subject should also be informed.

5.3.3 GDPR Rights of Data Subjects

In this section, there is a presentation of the GDPR rights of the MATRYCS data subjects. The relevant information that will be presented is explicitly described in Chapter III of the EU legislation [21] on the protection of natural persons with regard to the processing of personal data. More specifically, Article 12 summarises the rights of the data subjects, which will be analysed in this section. The main points are presented as follows:

- Right of information
- Right of access
- Right of rectification, erasure and blocking
- Right to object
- Right not to be subject to automated decisions

The **right to information**, which is presented in Articles 13 and 14, includes all the rights of the data subject to be sufficiently informed on the details of the data collection process, such as the contact details of the controller and the data protection officer, the purpose of the data processing, the people that may have access to the personal data etc. The right to information as described by EU law is presented below:

"Where personal data relating to a data subject are collected from the data subject, the controller shall, at the time when personal data are obtained, provide the data subject with all of the following information:

- The identity and the contact details of the controller and, where applicable, of the controller's representative;
- The contact details of the data protection officer, where applicable;
- The purposes of the processing for which the personal data are intended as well as the legal basis for the processing;

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

- Where the processing is based on point (f) of Article 6(1), the legitimate interests pursued by the controller or by a third party;
- The recipients or categories of recipients of the personal data, if any;
- Where applicable, the fact that the controller intends to transfer personal data to a third country or international organisation and the existence or absence of an adequacy decision by the Commission.

The **right of access** describes the right of the data subject to ask from the data processor information concerning the data that are being processed, as well as the right to demand that the results of the operations are communicated to him/her. In the case of MATRYCS, the LSPs will actively participate in the procedure and the individual results of each use case will be available to them in real time, thus satisfying the requirement of immediate access.

The **right of erasure**, analysed in Article 17, describes the right of the data subject to demand immediate delete of the collected personal information from the data controller in case that any of his/her rights are violated. The process of erasure could be temporary or definite. In case of a temporary erasure, the data processor could temporarily store the selected information on an alternate server or processing system, where these datasets would not be utilised. In case of a definite erasure, the data controller should remove all published data of the subject from any website or processing system. The temporary erasure of data is also referred as **right of blocking** because the data subjects' information is frozen and not eligible for any kind of operation. Finally, the data subject has the **right of rectification**, which could be also described as data correction and helps the data controller assure that the stored information is always in a correct form.

The **right to object** implies that initially none of the data subjects must be obliged to take part in the MATRYCS framework developing and research activities. Moreover, if a data subject that participates in the MATRYCS activities decides that he/she does not wish to continue participating by giving personal data, then he/she has the right to demand the termination of the usage of the collected personal data.

Finally, the **right not to be subject to automated decisions** which seriously affect the data controller or generate legal effects on the data controller is presented in Article 22 of the GDPR EU legislation. In the context of MATRYCS, this right means that the data of the pilots should not be utilised if there is a possibility of provoking any legal or ethical negative consequences for the pilots, such as discredit or calumniation.

As analysed in the above paragraphs all rights of the data subjects participating in the data collection procedures should be respected by MATRYCS, thus ensuring that the MATRYCS LSPs are supported in the greatest extent.

5.3.4 Data Protection Impact Assessment

The Data Protection Impact Assessment (DPIA) [24] is a procedure carried out in order to evaluate the origin, nature and severity of any possible risk concerning data protection. The results of this assessment are taken into consideration in the phase of the determination of the appropriate actions that must be performed in order to secure compliance with GDPR Regulation. The Data Protection Impact Assessment should incorporate for each risk described the severity of the risk, the likelihood of emergence of the risk, the measures that should be taken and the potential safeguards which can be considered, in order

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

to ensure the protection of personal data.

According to EU legislation [21], *"A Data Protection Impact Assessment is equally required for monitoring publicly accessible areas on a large scale, especially when using optic-electronic devices or for any other operations where the competent supervisory authority considers that the processing is likely to result in a high risk to the rights and freedoms of data subjects, in particular because they prevent data subjects from exercising a right or using a service or a contract, or because they are carried out systematically on a large scale"*. Therefore, the need for development of a Data Protection Impact Assessment in the context of the MATRYCS project should be considered in the following months according to the final datasets which will be used during the development phase.

According to Article 35 of the GDPR EU Legislation the minimum context of a Data Protection Impact Assessment should include:

- A description of the processing and its purposes.
- An assessment of the necessity and proportionality of the processing in relation to the purpose of the processing.
- An assessment of the risks to the rights and freedoms of data subjects.
- The measures envisaged to address the risks, including security measures and mechanisms, to ensure the protection of personal data and to demonstrate compliance with the GDPR.

5.3.5 Methodology for personal data processing

In the context of Task 2.4 Regulatory Framework for Data Protection, IPR and Ethical Issues, concerning the GDPR requirements of the MATRYCS project associated activities, a Questionnaire (Figure 9) concerning the GDPR Issues was developed and assigned to all LSPs. The necessity of this Questionnaire stems from the fact that the LSPs are sharing specific internal data which will be used for the development of the MATRYCS framework. Thus, it is essential to monitor the origin of the shared data, as well as to detect whether any personal or sensitive data will be shared and used. The MATRYCS GDPR Questionnaire along with the answers of the pilots are presented in the Appendix of this deliverable.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

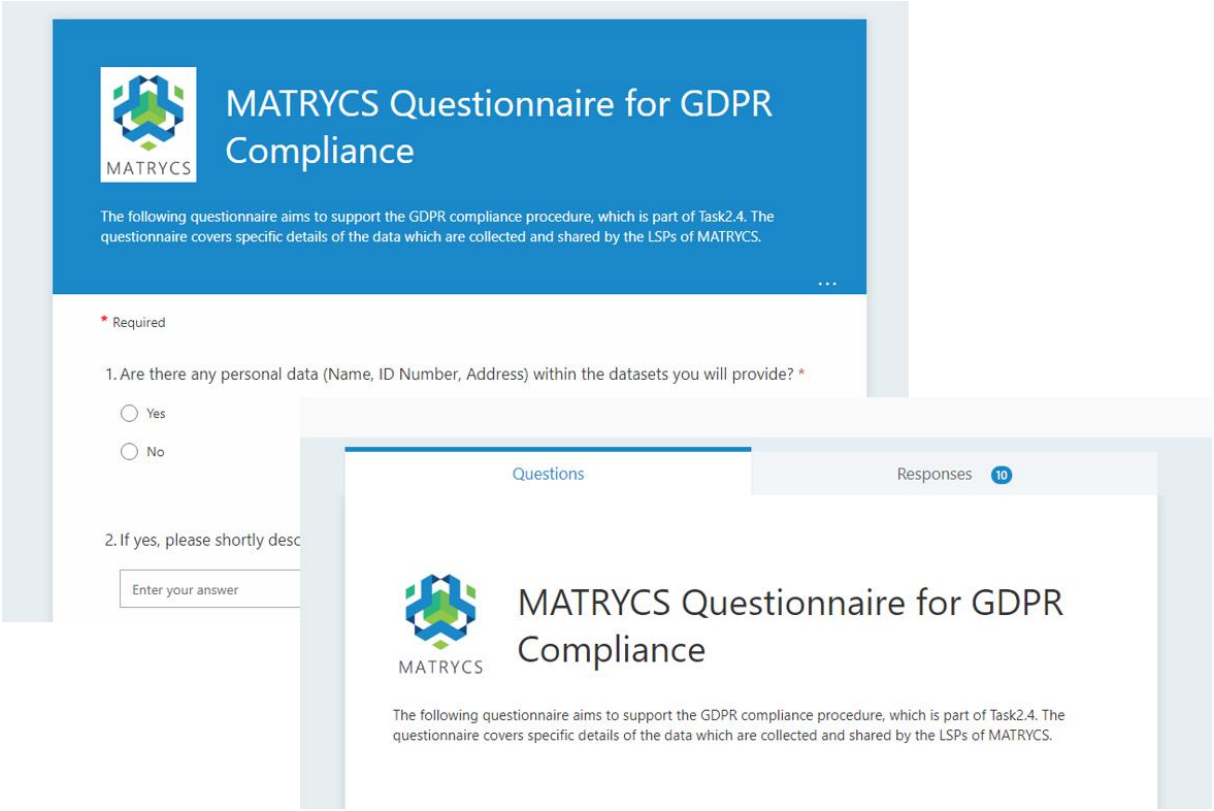


Figure 9: MATRYCS Questionnaire for GDPR Issues Compliance

As a result, the following technical and organisational measures will be implemented to ensure a high level of security:

- Anonymisation & pseudoanonymisation of datasets with personal or sensitive information
- Data recovery mechanism which will secure data availability and access to personal data for the data subject at any time, even in case of a technical or physical incident
- High confidentiality and integrity when processing sensitive data during the test phase and resilience of processing systems during the deployment phase
- Regular testing processes in order to assess the effectiveness of the security mechanisms that will be developed

All these measures will take effect during the development phase of the MATRYCS framework and will be designed according to the likelihood and the severity of harm in case of personal data breach.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

6 Future Steps

In deliverable a first version of user requirements and technical and specifications was presented. Additionally, security requirements were also addressed, analysing the IPR and the Ethical Issues of the MATRYCS project, the Network and Information Security requirements and the GDPR regulatory framework. All actions taken for the specification elicitation reported and mapped within the BDVC of MATRYCS framework so as to showcase the connection between each activity.

A concrete set of 105 user requirements were presented which led in 60 technical specification in total. More specifically, the user requirements were categorised in 74 functional requirements and 31 non-functional requirements. On the other hand, the technical requirements were mapped to the three MATRYCS layers, resulting in 20 MATRYCS GOVERNANCE requirements, 16 MATRYCS PROCESSING requirements and 16 MATRYCS ANALYTICS requirements. Finally, 8 general requirements were, also, extracted concerning the overall functionalities of the MATRYCS framework.

The activities presented in this deliverable will be in progress until M18 of the project and will be revised in parallel with the development of the MATRYCS components. Therefore, depending on the needs that will arise up to M18, User Requirements and Technical and Security Specifications can be modified. Their final version will be delivered in D2.4 MATRYCS Reference Architecture for Buildings Data which is due in M18 and after the first MATRYCS framework technology releases.

The work reported in this deliverable allows the preparation of the definition and delivery of the open Reference Architecture for Buildings Data, which aims to satisfy the described specifications, to ensure compatibility with existing dataset formats and to allow integration with existing architectures.

Regarding security specifications, the next steps involve the continuous monitoring of the collected datasets in order to identify the processed personal and sensitive data and guarantee the secure handling of this information during the development phase of the project.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

References

[1] E. Curry, "The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches," *New Horizons for a Data-Driven Economy,* pp. 29-37, 2016.

[2] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. Hung Byers, Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute, 2011.

[3] D. Laney, "3D data management: Controlling data volume, velocity and variety.," 2001.

[4] A. Jacobs, "The pathologies of big data," *Communications of the ACM,* pp. 36-44, 2009.

[5] M. E. Porter, "Creating and sustaining superior performance," *Competitive advantage,* pp. 167-206, 1985.

[6] D. Connect, "A European strategy on the data value chain," 2013.

[7] X. Feng, A. Kumar, B. Recht and C. Ré, "Towards a unified architecture for in-RDBMS analytics," in *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, 2012.

[8] V. Marinakis, H. Doukas, J. Tsapelas, S. Mouzakitis, Á. Sicilia, L. Madrazo and S. Sgouridis, "From big data to smart energy services: An application for intelligent energy management," *Future Generation Computer Systems,* vol. 110, pp. 572-586, 2018.

[9] U. Berardi, "Building Energy Consumption in US, EU, and BRIC Countries," *Procedia Engineering,* no. 118 , pp. 128-136, 2015.

[10] V. Marinakis, "Big data for energy management and energy-efficient buildings," *Energies,* vol. 13, no. 7, p. 1555, 2020.

[11] N. oseleva and G. Ropaite, "Big data in building energy efficiency: understanding of big data and main challenges," *Procedia Engineering,* vol. 172, pp. 544-549, 2017.

[12] V. Marinakis, H. Doukas, C. Karakosta and J. Psarras, "An integrated system for buildings' energy-efficient automation: Application in the tertiary sector," *Applied Energy,* vol. 101, pp. 6-14, 2013.

[13] V. Marinakis and H. Doukas, "An advanced IoT-based system for intelligent energy management in buildings," *Sensors,* vol. 18, p. 610, 2018.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

[14] "Designing Buildings," [Online]. Available: https://www.designingbuildings.co.uk/wiki/Big_data_for_buildings.

[15] M. Broy, "Multifunctional software systems: Structured modeling and specification of functional requirements," *Science of Computer Programming,* vol. 75, no. 12, pp. 1193-1214, 2010.

[16] L. Chung and J. C. S. do Prado Leite, "On non-functional requirements in software engineering," *Conceptual modeling: Foundations and applications* , pp. 363-379, 2009.

[17] "JamaSoftware," [Online]. Available: https://www.jamasoftware.com/blog/characteristics-of-excellent-requirements/.

[18] "COM(2016) 410 final," [Online]. Available: https://ec.europa.eu/transparency/regdoc/rep/1/2016/EN/1-2016-410-EN-F1-1.PDF.

[19] "EUR-Lex Access to European Union law," [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.194.01.0001.01.ENG&toc=OJ:L:2016:194:TOC.

[20] "Directive (EU) 2016/1148 of The European Parliament and of The Council of 6 July 2016, concerning measures for a high common level of security of network and information systems across the Union.," no. https://eur-lex.europa.eu/legal-content.

[21] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC," [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[22] "GENERAL DATA PROTECTION REGULATION (GDPR)," [Online]. Available: https://gdpr-info.eu/art-5-gdpr/.

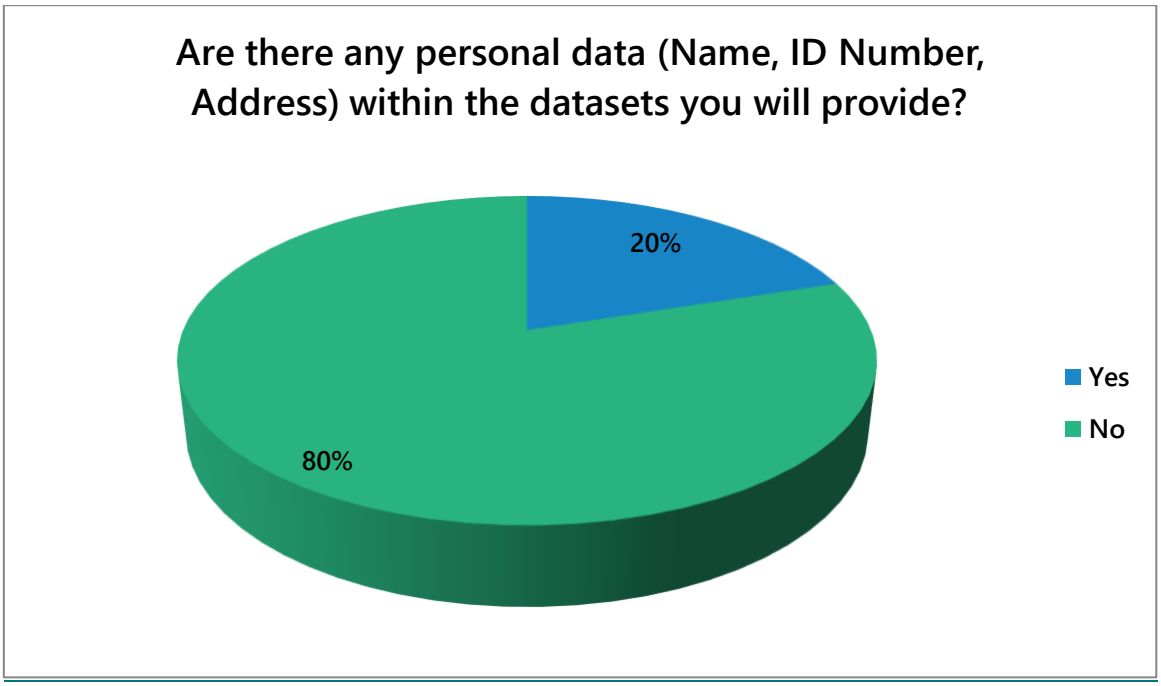
[23] "GENERAL DATA PROTECTION REGULATION (GDPR)," [Online]. Available: https://gdpr-info.eu/art-6-gdpr/.

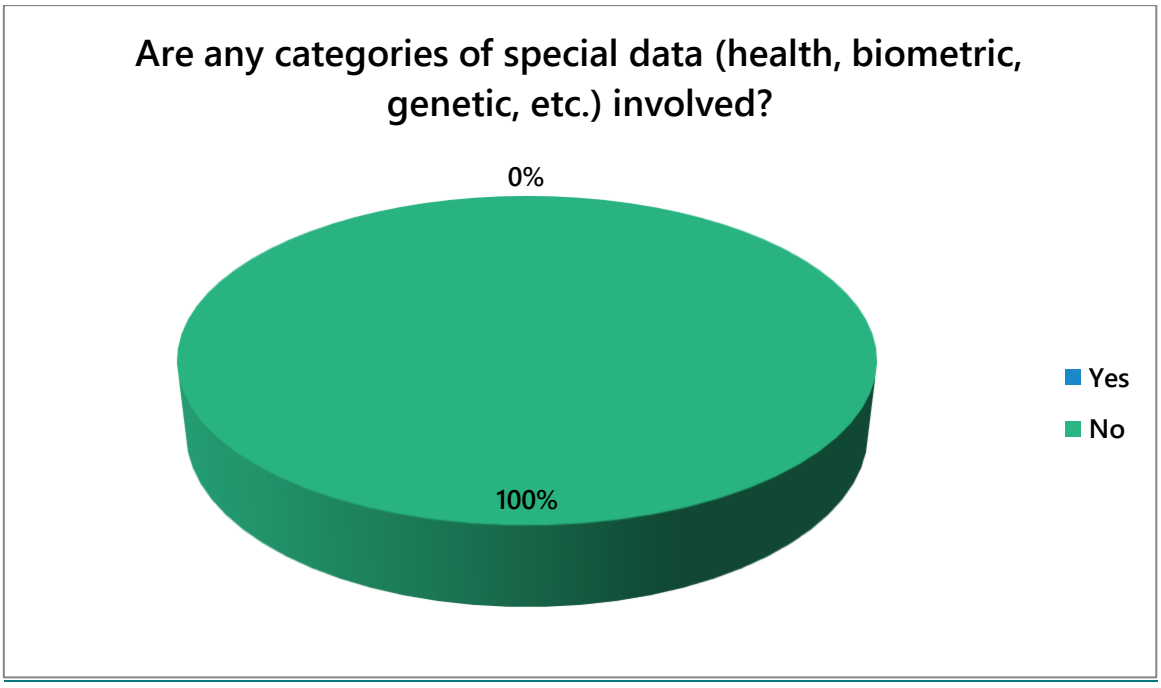
[24] "GENERAL DATA PROTECTION REGULATION (GDPR)," [Online]. Available: https://gdpr.eu/data-protection-impact-assessment-template/.

The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

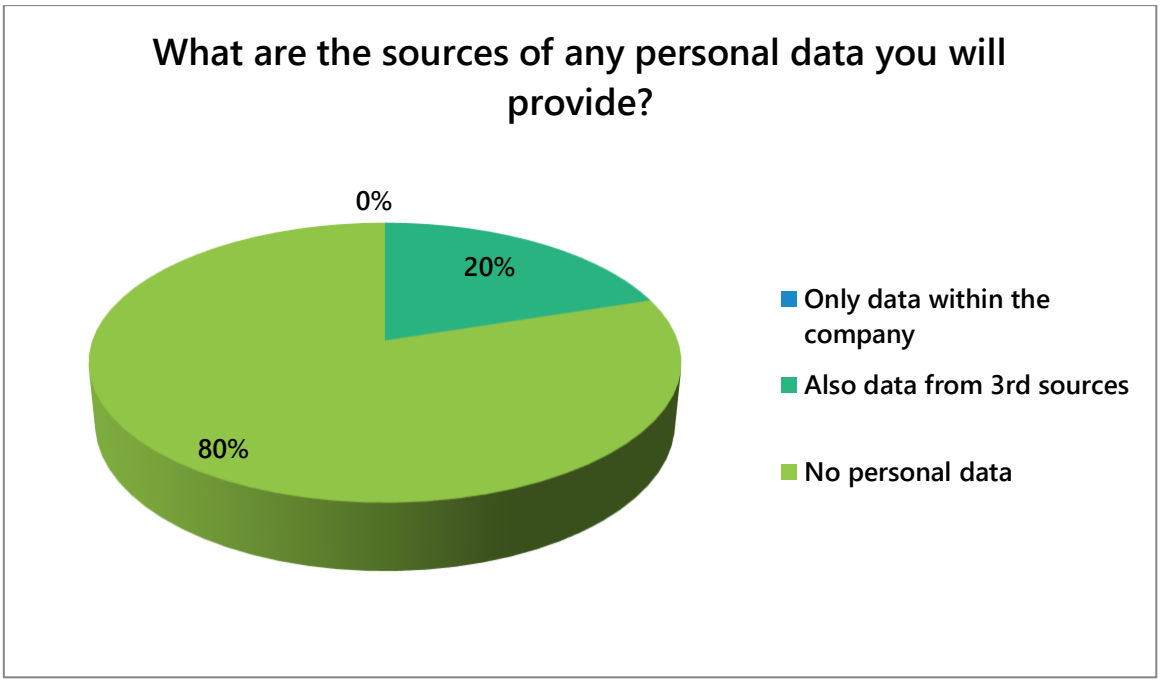
Appendix

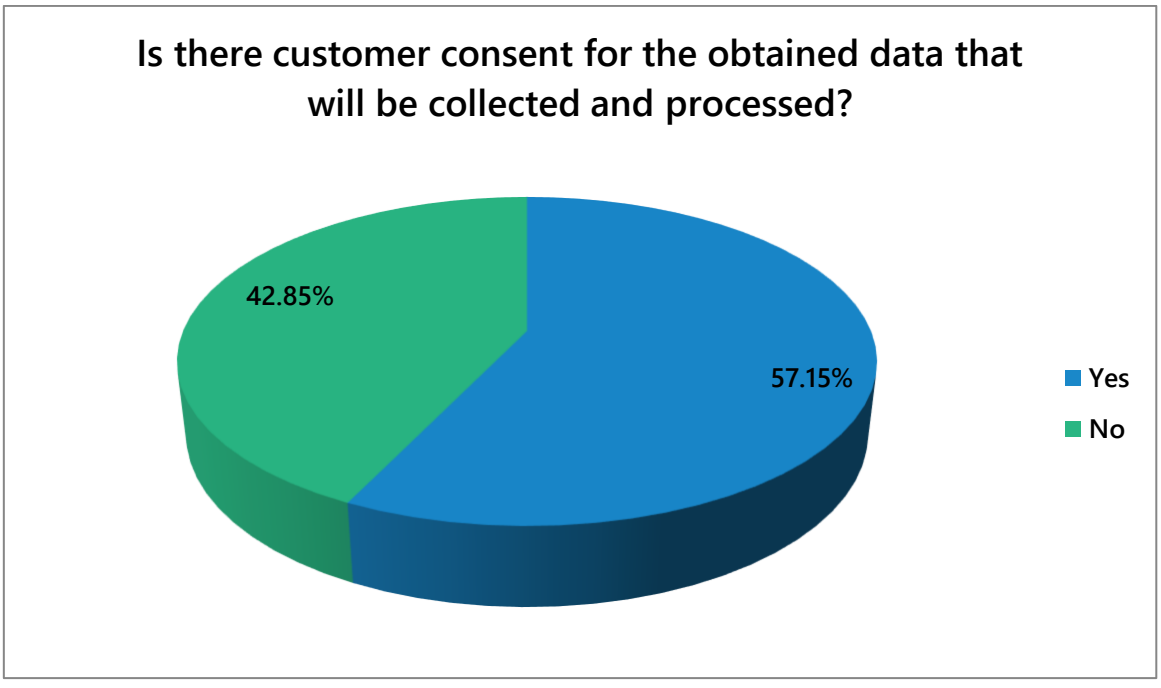
This Appendix includes the results of the Questionnaire for GDPR Issues, which was developed in the context of Task 2.4 Regulatory Framework for Data Protection, IPR and Ethical Issues, and more specifically the activity regarding compliance to GDPR legislation.



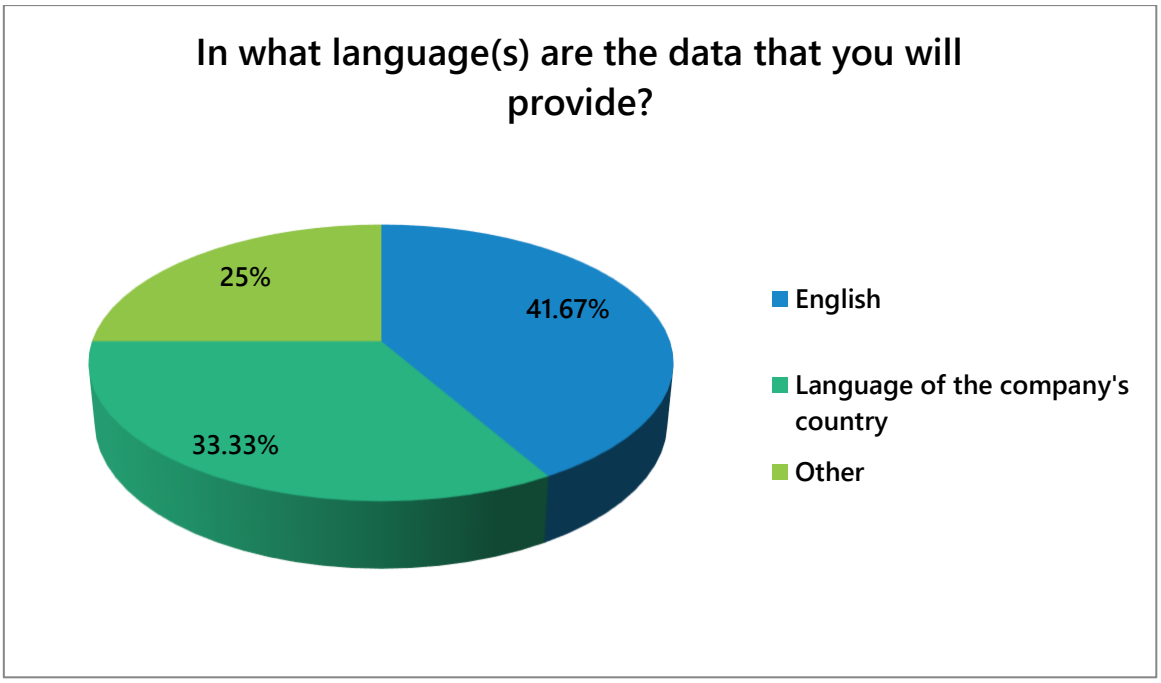


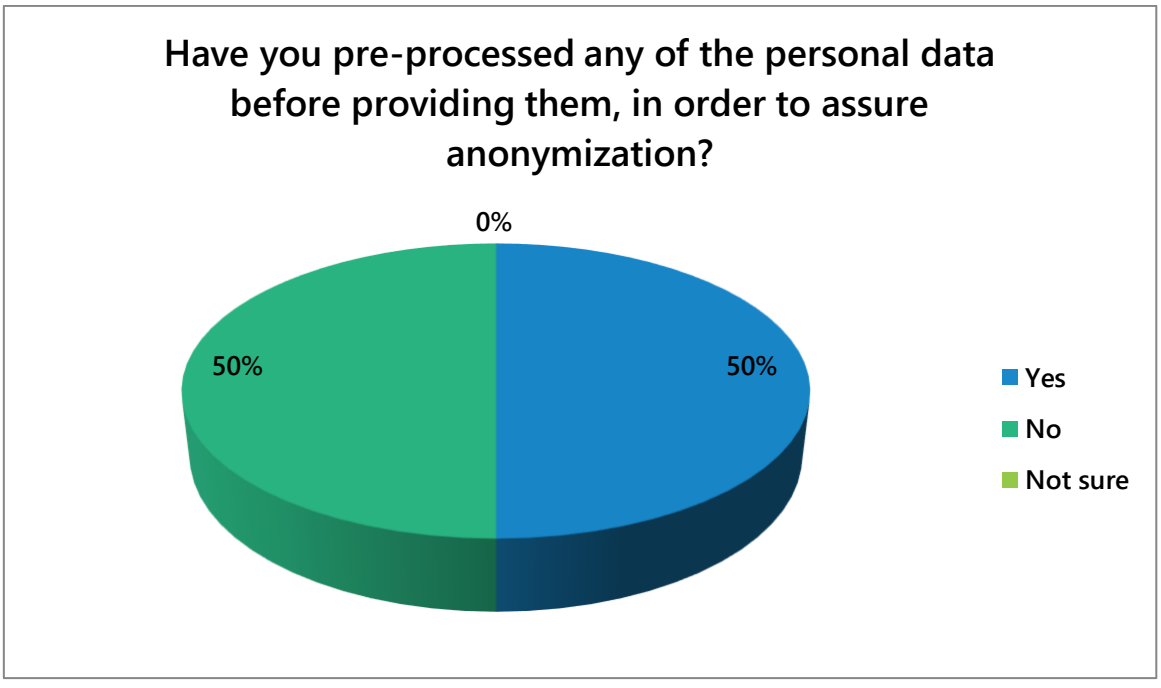
The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158



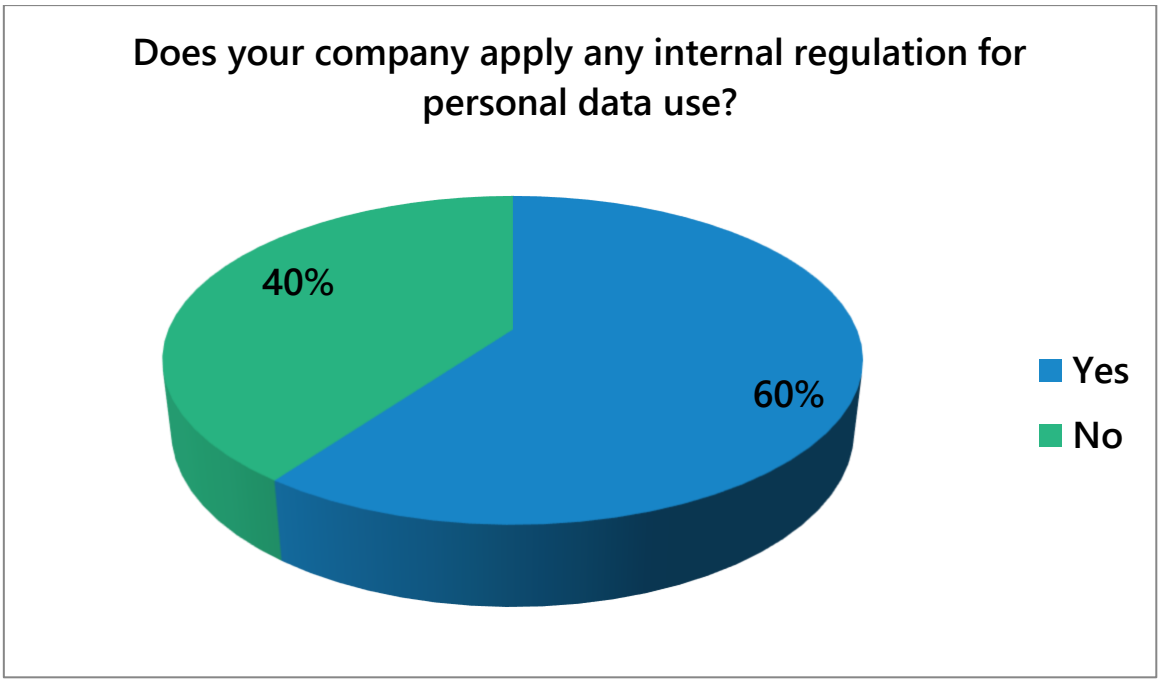


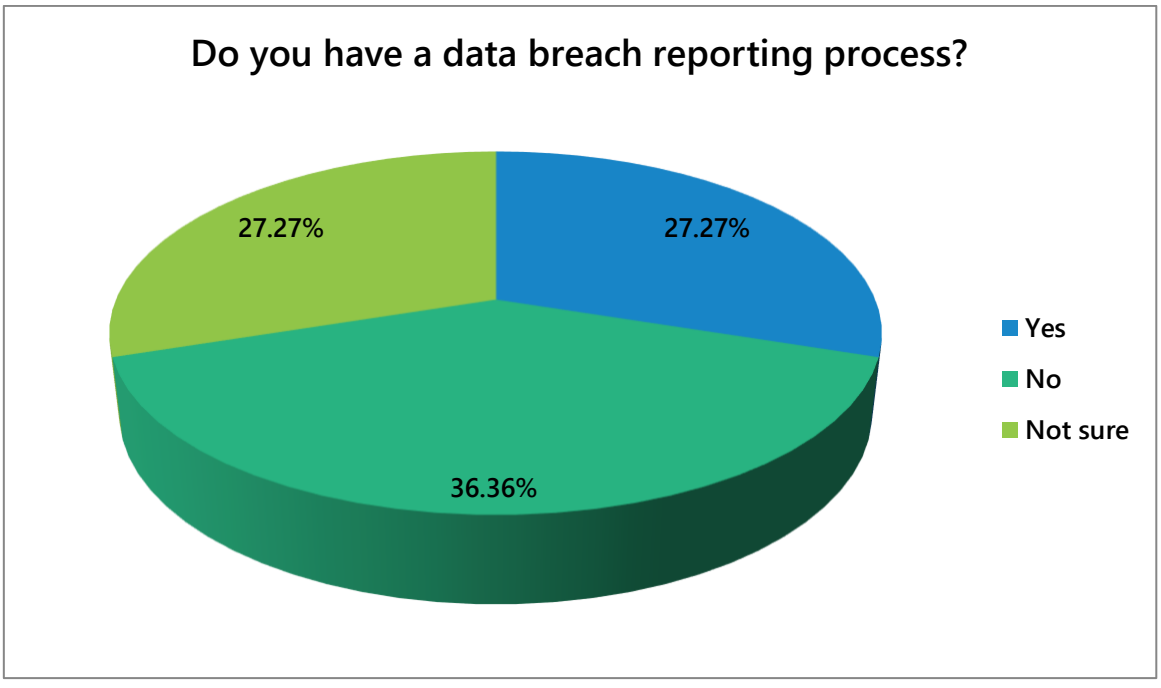
The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158



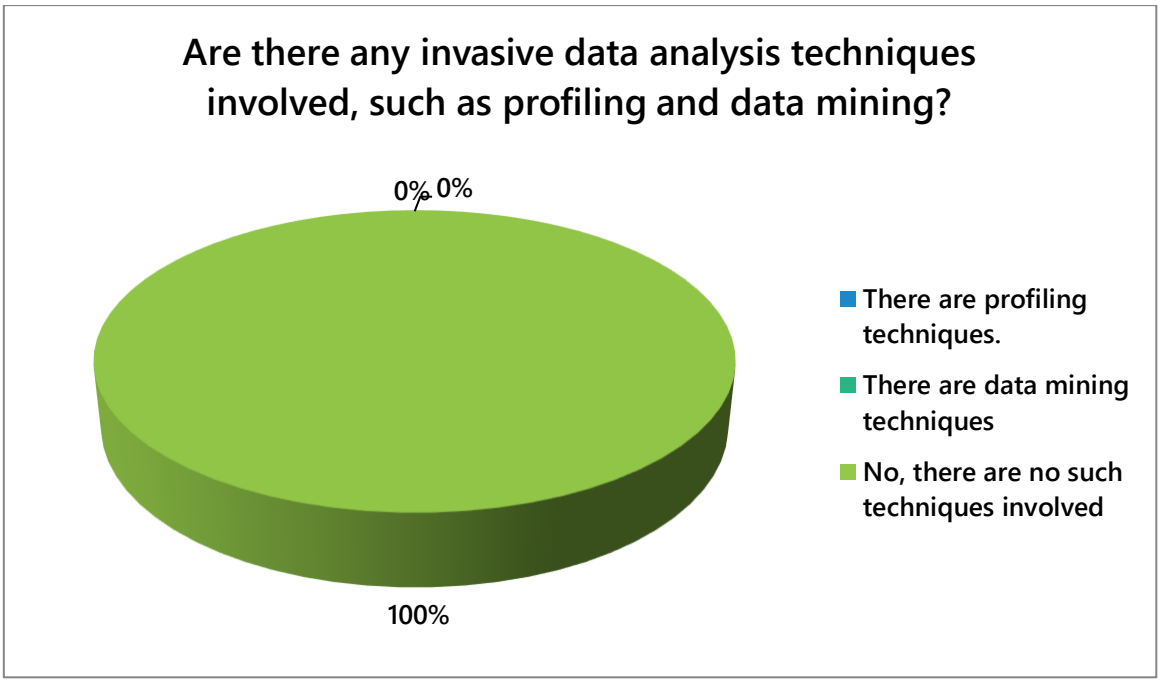


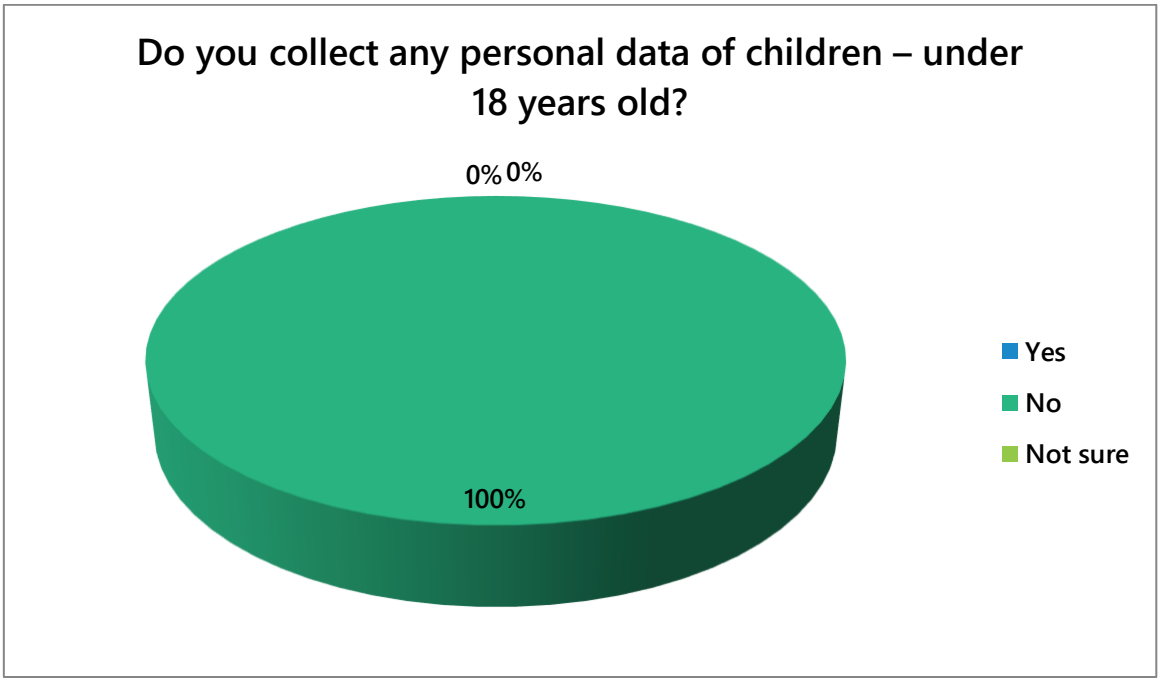
The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158



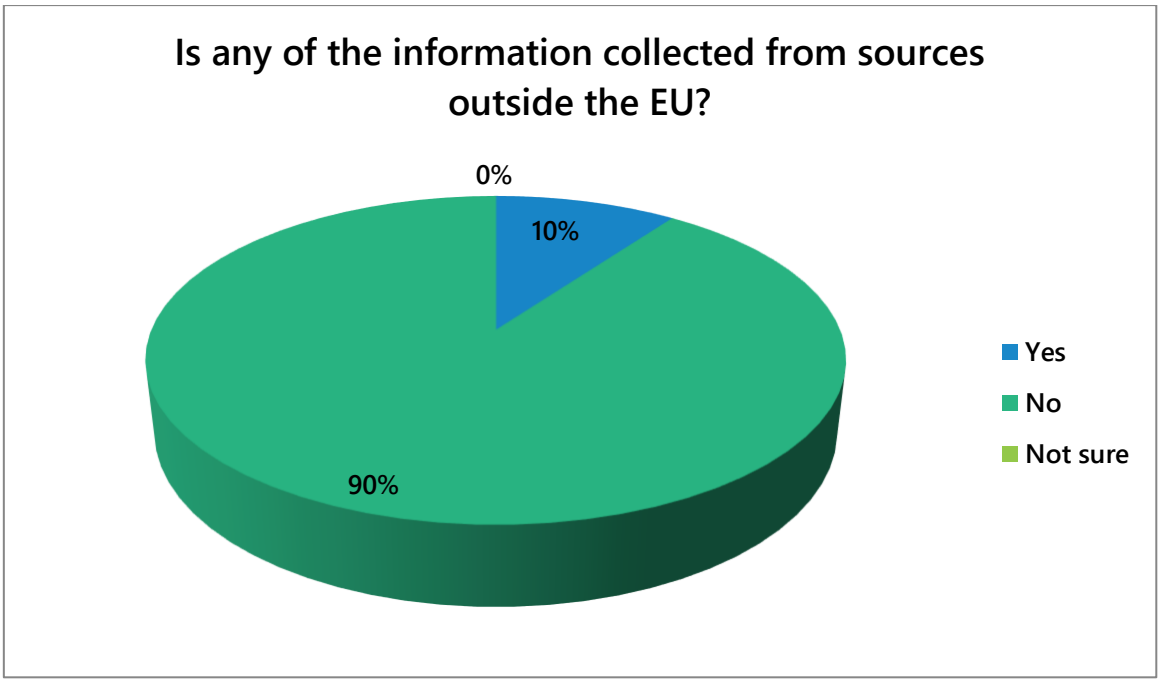


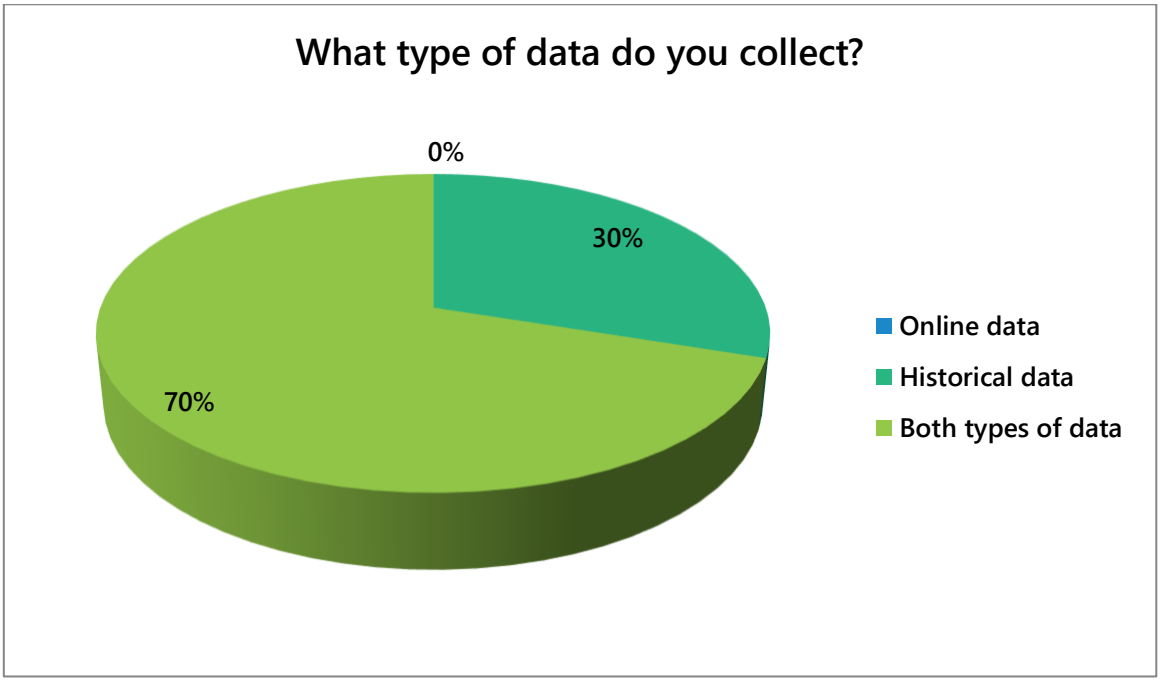
The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158



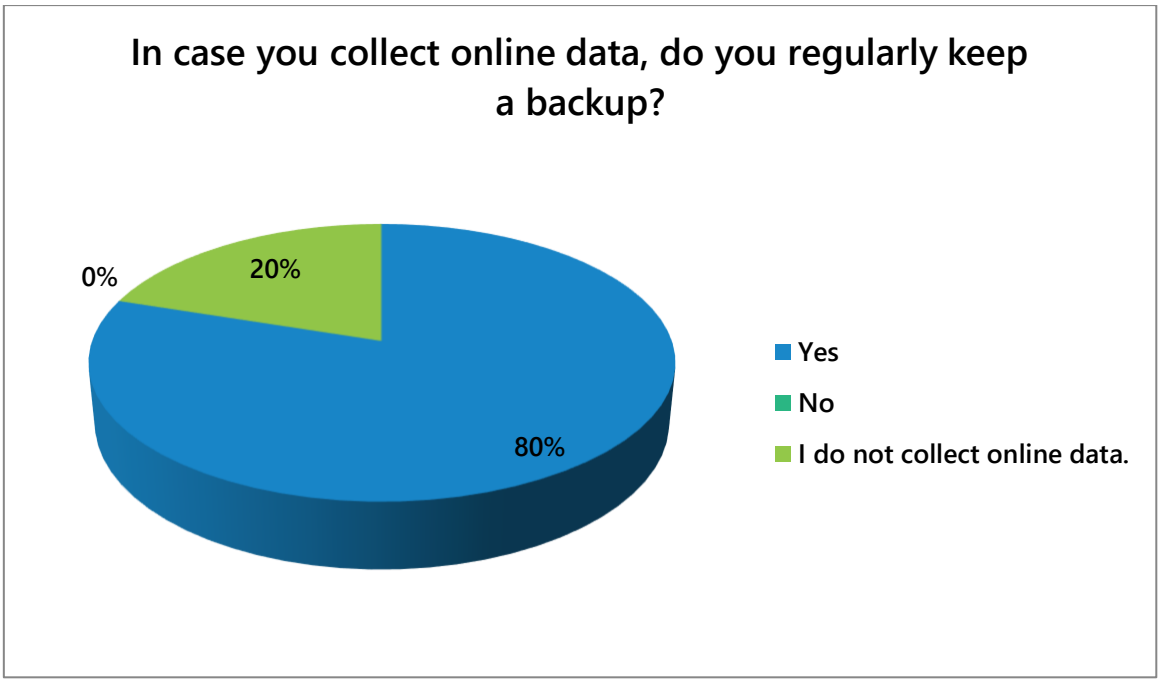


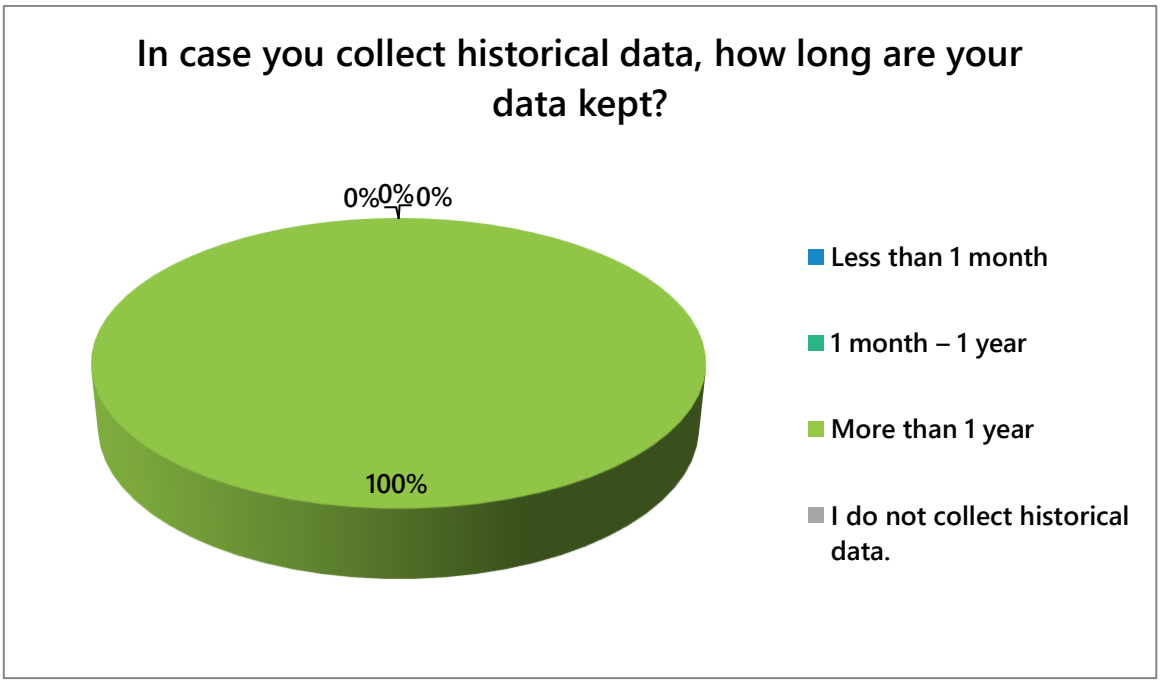
The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158



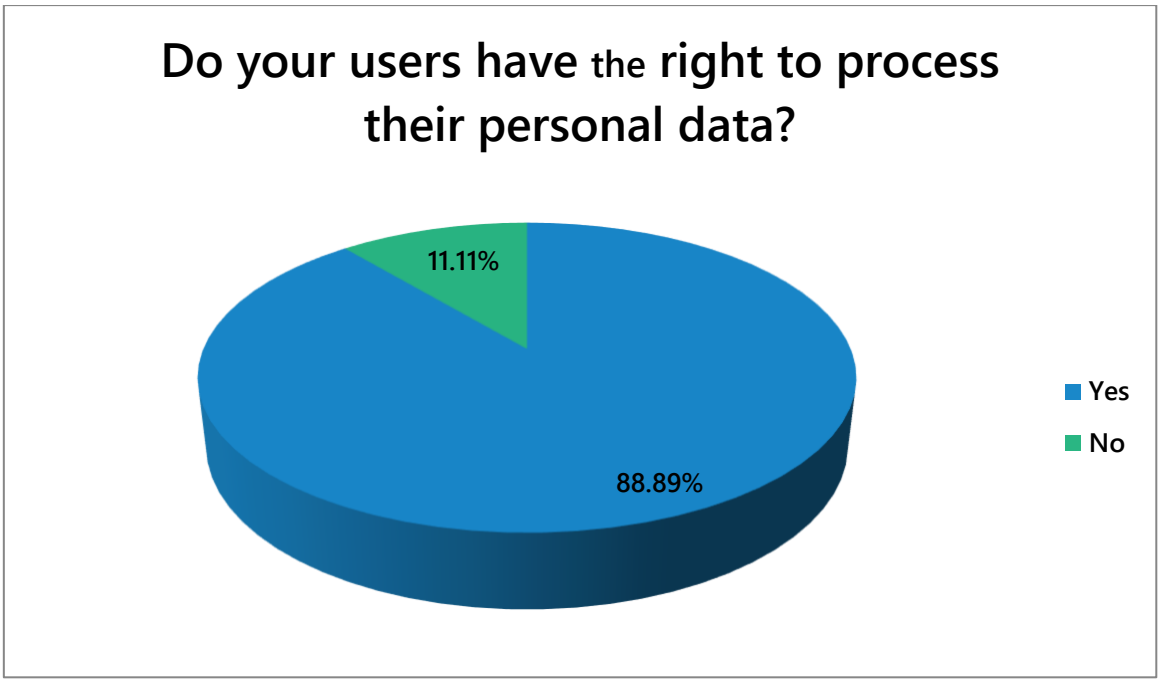


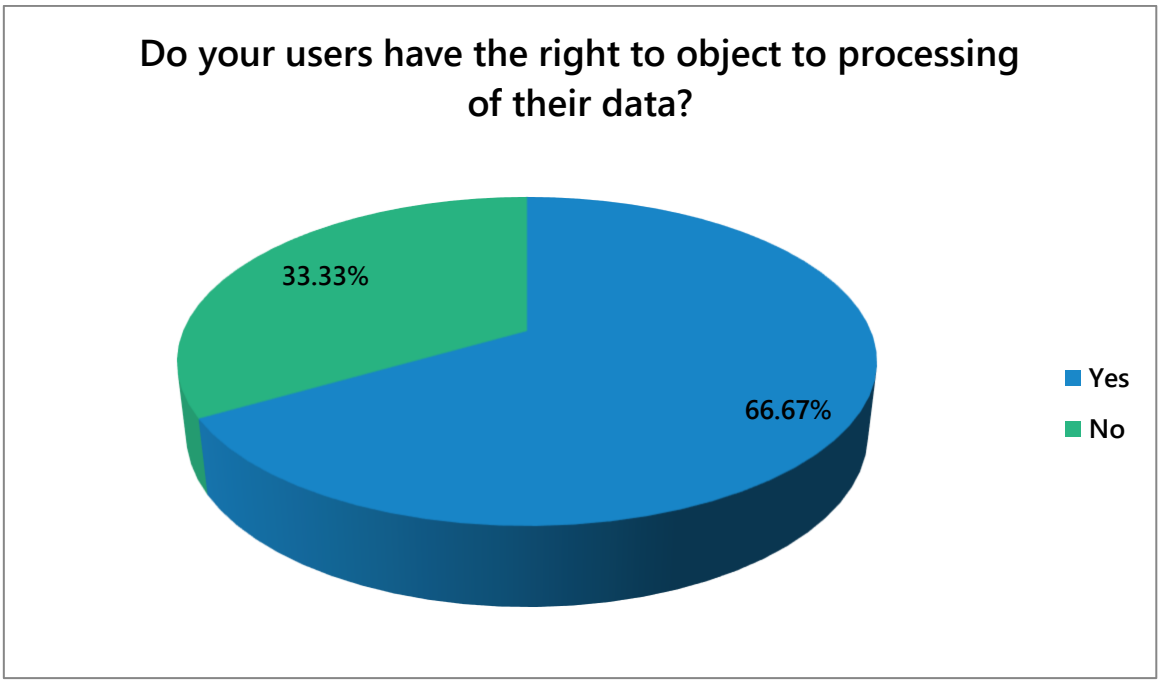
The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158



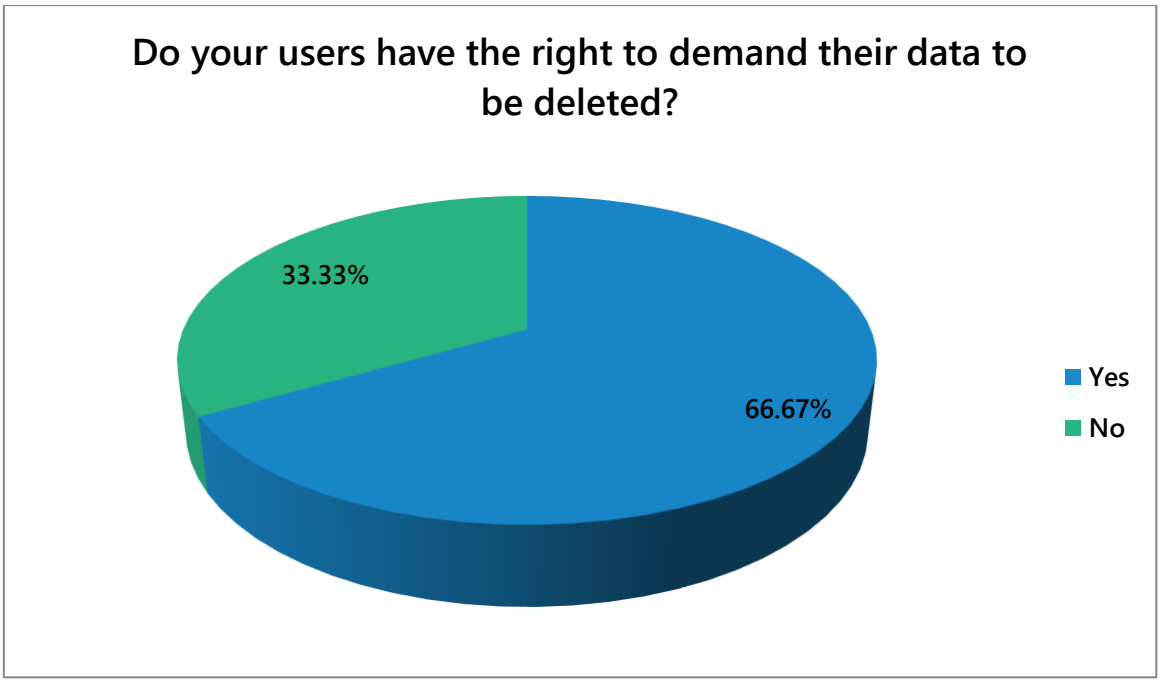


The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158





The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158



The MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158