MATRYCS

Deliverable 2.4

MATRYCS Reference Architecture for Buildings Data (final version)

WP2 – System Requirements and Specifications

March 2022

www.matrycs.eu

Modular Big Data Applications for Holistic Energy Services in Buildings

MATRYCS

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	D2.4 – MATRYCS Reference Architecture for Building Data (final version)		
Work Package	WP2 – System Requirements and Specifications		
Delivery Month (DoA)	March 2022	Version	1.0
Actual Delivery Date	<01/04/2022>		
Nature	Report	Dissemination Level	Public
Lead Beneficiary	RWTH Aachen University		
Authors	Zhiyu Pan [RWTH], Stefan Lankes [RWTH], Dario Pellegrino [ENG], Pasquale Andriani [ENG], Leandro Lombardo [ENG], Francesco Saverio Nucci [ENG], Panagiotis Kapsalis [NTUA], Konstantinos Alexakis [NTUA], Elissaios Sarmas [NTUA], Nikos Dimitropoulos [NTUA], Zoi Mylona [HOLISTIC], Daniele Antonucci [EURAC], Gregor Cerar [COMSENSUS], Miha Smolnikar [COMSENSUS], Sofía Mulero Palencia [CARTIF], Gema Hernández Moral [CARTIF], Carla Rodríguez Alonso [CARTIF]		
Quality Reviewer(s):	Roman Mendle [ICLEI], Vangelis Marinakis [NTUA]		
Keywords	Open Reference Architecture, Technical Specifications and Requirements, Big Data Management, Data Governance, Data Processing, Building Services, Big Data Analytics, Semantic Interoperability, Big Data Value Chain, Data Spaces		

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

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 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

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

Tables

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

Abbreviation and Acronyms

Acronym	Description
AI	Artificial Intelligence
AIOTI	Alliance for Internet of Things Innovation
API	Application Programme Interface
BDVA	Big Data Value Association
BEMS	Building Energy Management System
BIM	Building Information Model
BSP	Building Service Provider
BVC	Building Value Chain
DaaS	Data as a Service
DAP	Data Analytics Provider
DL	Deep Learning
EPC	Energy Performance Certificate
GDPR	General Data Protection Regulation
GE	Generic Enabler
HLA	High Level Architecture
HPC	High Performance Computing
IaaS	Infrastructure as a Service
ICT	Information and Communication Technology
IDS	International Data Space
IDSA	International Data Space Association
IIC	Industrial Internet Consortium
IIRA	Industrial Internet Reference Architecture

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

Acronym	Description
IoT	Internet of Things
IPR	Intellectual Property Right
IT	Information Technology
ML	Machine Learning
PaaS	Platform as a Service
PPP	Public-Private Partnership
QoS	Quality of Service
RAF	Reference Architecture Framework
RAM	Reference Architecture Model
RDF	Resource Description Framework
SaaS	Software as a Service
SME	Small-Medium Enterprise
SP	Service Provider
TP	Technology Provider
WG	Working Group

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 is the final version of the MATRYCS Reference Architecture. In the first version of the MATRYCS Reference Architecture, a review of already existing works aimed at the definition of reference architectural models is firstly presented. In the final version, two more reference architectural models are analysed. The main aspects and design principles are considered as additional input for the definition of the MATRYCS Reference Architecture. The specification, technical and non-technical requirements for the MATRYCS Reference Architecture have been described in the previous Deliverable. In this Deliverable, the refined specification, technical and non-technical requirement extracted from the use cases are updated. They are recalled here to highlight how they are fulfilled via the different architecture layers or components. Given the broad scope and the different focus of the Large-Scale Pilots in the MATRYCS project, their use cases provide a heterogeneous set of requirements. Considering them has the goal to ensure that the proposed Reference Architecture fits the specific requirements possibly existing in the building domain.

In the next part of the Deliverable, the functional description of the different layers and components of the proposed Reference Architecture are presented. The adopted architecture is composed of three different layers: MATRYCS governance layer, MATRYCS processing layer, MATRYCS analytics toolbox layer. Additionally, two different views of MATRYCS Reference Architecture are designed, which aims to show two perspectives of the Reference Architecture: functional view and technology view.

The Deliverable will provide the overall view of these layers and of the related components, specifying how they interact each other and which technical and non-technical functionalities they must provide. Finally, the Deliverable presents the alignment of the MATRYCS Reference Architecture with the three additional reference architectural models presented, in order to highlight existing analogies, possible differences or gaps, and to demonstrate how the proposed architecture can be mapped to different architectural views. A mapping of the MATRYCS Reference Architecture components into the building blocks of a future data space for buildings is finally presented.

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 document

Sensors, smart meters, Internet of Things (IoT) devices, and a variety of other data sources are contributing to the digitalization and smartification of the European construction industry, which is resulting in the collection of a massive quantity of data. Exploiting this data and harnessing their value through ad hoc analytics services is critical for increasing energy efficiency in the building sector and progressing toward the European Commission's decarbonisation and emission reduction objectives. Unfortunately, relevant data is currently fragmented and isolated in non-interoperable data silos, making it difficult for parties engaged in the design and provision of building services to access it. Furthermore, building data is derived from a variety of heterogeneous data sources and is characterized by a variety of data formats, sizes, resolutions, granularity levels, truthfulness, and other characteristics, all of which impede the efficient use of data. Overall, this makes extracting value from these data and developing enhanced services that may speed up the shift to smart energy-aware buildings exceedingly challenging. Given this scenario, one of the main objectives of the MATRYCS project is to define and deploy a Reference Architecture for building data exchange, management and real-time processing. The goal of this Deliverable is to present and discuss the final version of the open Reference Architecture for Building Data conceived within the MATRYCS project. In particular, this Deliverable presents the concepts, principles and requirements taken into account for the definition of the proposed Reference Architecture, which update the first version of reference architecture.

1.2 Positioning within the project

This Deliverable is the final outcome of the activities in the Task T2.5 "Big Data, AI and IoT Reference Architecture and Alignment with Existing Frameworks and Architectures", which is the final task of the WP2 "System Requirements and Specifications".

As a matter of fact, Task 2.5 (and consequently this Deliverable and D2.3) directly or indirectly uses as input the results of all the other tasks of WP2. In particular, the Reference Architecture here presented looks at the specifications and requirements identified in T2.3 ("Analytics Building Services Specifications") and T2.4 ("Regulatory Framework for Data Protection, IPR and Ethical Issues"), which have been presented in Deliverable D2.2. Specifications, technical and non-technical requirements have been extracted from the analysis of the MATRYCS use cases, which were defined as part of the activities in T2.2 ("User Stories and Requirements Analysis"). Therefore, the D2.4 takes over all the latest update from all those tasks and the input from D2.3.

The activities and the Reference Architecture presented in this Deliverable represent an input for the different tasks of WP3, WP4 and WP5, which are all dedicated to the development of the specific modules, tools and services included in the three layers of the considered architecture (respectively, the MATRYCS governance layer, the MATRYCS processing layer and the MATRYCS analytics layer). It is worth noting that, since the proposed architecture provides a high-level abstraction of the required functionalities, it does not put any constraints on the technologies used for the practical

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

implementation of the architecture components; the implementation work done in WP3, WP4 and WP5 (reported in the associated Deliverables) thus represents only a particular instantiation of the presented reference architecture.

1.3 Deliverable structure

In addition to the present Section, this Deliverable is composed of other four Sections:

- **Section 2** provides a review of two already existing works towards the definition of reference architectural models for big data management, which is the same structure as in D2.3.
- **Section 3** focuses on the updating the specifications and requirements coming from different use cases associated to operational performance, retrofitting, policies and investments in the building sector. The Security and Privacy requirement are discussed in more detail.
- **Section 4** presents and discusses the final version of the proposed MATRYCS Reference Architecture, which two different views have been designed. It will provide the details on the different architecture layers, the associated modules and components, together with their functional description and the requirement met.
- **Section 5** discusses the alignment of the proposed Reference Architecture to three additional existing frameworks for big data management already presented in Section 2 and D2.3, with the goal of highlighting similarities, possible differences and to show how the proposed architecture can be mapped into different architectural views.

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

2 Review of existing architectural models for big data management

This Section provides an additional review of some of the most important proposals of reference architecture models for big data management compared with D2.3. The scope is to underline the major aspects that should be considered for the design of a reference architecture as well as to identify how the proposed MATRYCS architecture can be aligned and mapped to already existing architectures. The design principles and reference models are described in D2.3 as following:

- the Big Data Value Association
- the Industrial Internet Consortium
- the Alliance for Internet of Things Innovation Business Processes
- the FIWARE Community
- the International Data Space Association
- the OpenDEI project

In the following, the GAIA-X project [1], BRIDGE [2] will be presented and discussed.

2.1 GAIA-X Reference Architecture

GAIA-X is a European project that aims to create an Infrastructure and Data Ecosystem based on European ideals and standards. Its architecture is guided by this ultimate aim, which is to allow interaction amongst all actors in the European digital economy, utilising digital processes and information technology. It provides open, consistent, quality-assured, and easy-to-use creative data interchange and services by using existing standards, open technology, and concepts. A common and stable base made up of Policy Rules and an Architecture of Interconnection Standards supports the GAIA-X ecosystem. The GAIA-X architecture is depicted in Figure 1 in a high-level overview. GAIA-X specifies technological ideas, federation and interoperability capability (such as Identity and Access Management) that apply to the GAIA-X ecosystem as a whole. GAIA-X takes on the role of conductor, however, is not involved in individual participant transactions. Instead, GAIA-X allows Providers to improve their existing isolated solutions by making them GAIA-X-enabled. A collection of Federation Services is created, implemented, and operated to bring the architectural principles to life. The phrase Federation Service refers to both infrastructure and organizational support functions including onboarding and certification.

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

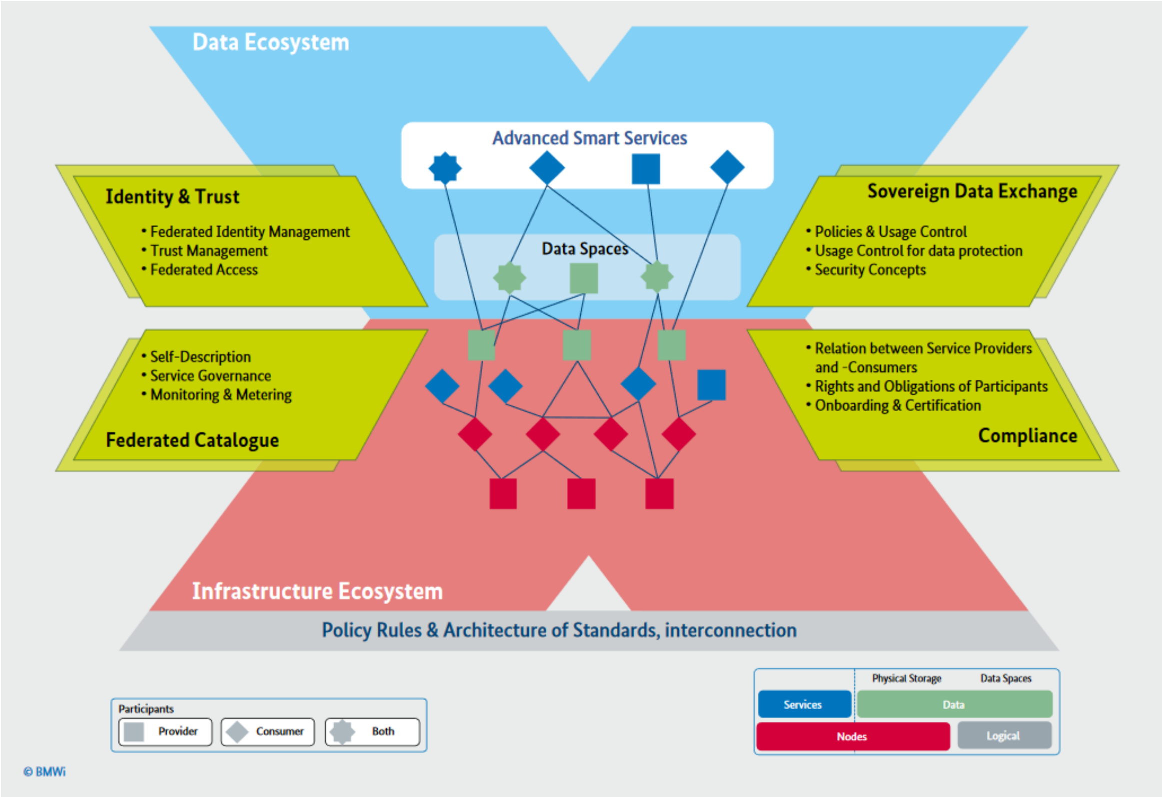


Figure 1: GAIA-X Reference Architecture Framework

Federation Services are grouped into four domains:

- **Identity and Trust:** Identity and trust are viewed from several perspectives across the architectural stack.
  - Federated Identity Management: describing the provisioning of identifiers also used for authentication.
  - Trust Management: aiming at establishing trust for every GAIA-X interaction.
  - Federated Access: specifying how access can be managed in a federated fashion.
- **Federated Catalogue (Interoperability):** The Catalogue comprises the GAIA-X ecosystem's Providers' offers. Different perspectives on current offers, such as domain-specific Asset choices, might be represented via a Catalogue. This functionality will be useful as long as the Consumer is aware of and capable of easily turning off any catalogue limitation and gaining access to all available Assets.
  - Self-Description: GAIA-X Self-Descriptions Express Asset and Participant attributes. A GAIA-X Self-Description describes an Asset's or Participant's properties and claims. Self-Descriptions are linked to the corresponding Asset or Participant's Identifier.
  - Service Governance: It define the relation between service provider and consumer.
  - Monitoring and Metering: Monitoring is a crucial feature of federated systems, especially cloud services. The loose coupling of Services and Nodes from various Providers is a technical challenge due to a diverse technological environment.

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

- **Sovereign Data Exchange:** Usage control techniques and an overarching security concept assure data interchange sovereignty. In addition, standards for data interchange compatibility will be chosen.
  - Policies and Usage Control: Policies in GAIA-X specify a set of constraints. They can be thought of as the opposite of the Self-Description.
  - Usage Control for data protection: GAIA-data X's protection approach is general, and it may be used to assure compliance with any data security standard.
  - Security Concepts: GAIA-mission X's is to expedite and broaden the adoption of safe and trustworthy data services. The data services will be housed in an open environment that enables for secure and compliant data exchange and processing across many parties in a secure and compliant manner. GAIA-X provides rules, regulations, and a target architecture to assure the greatest degree of data protection, security, transparency, and portability for all services, as well as the technical Federation Services that must be implemented by GAIA-X Participants.
- **Compliance:** Security and Data Protection depend not only on technical solutions, but also on organizational and governance aspects.
  - Relation between Service Providers and Consumers: Between two or more Participants, a Legal Context exists. Legal contexts aren't often expressed openly in technical systems. Consumption of services takes place within a Service Context, which is where Policies reside. A Legal Context includes Service Contexts. One of the participants must be the Service Instance Provider.
  - Rights and Obligations of Participants: Rights and responsibilities of each Participant are defined in the GAIA-X architecture.
  - Onboarding and Certification: The Provider must first register on GAIA-X before offering Services and Nodes. The Self-Description supplied by the Provider applying to integrate Services/Nodes in the GAIA-X environment is an important foundation for the onboarding process. The Provider should fill this Self-Description using a tool made accessible through the GAIA-X portal and APIs.

2.2 BRIDGE Reference Architecture

BRIDGE is a European Commission initiative in which 64 Horizon 2020 projects in the areas of Smart Grid, Energy Storage, Islands, and Digitalisation are participating to pool field experience, feedback, and lessons learned in order to create a structured view of cross-cutting issues that are encountered in the demonstration projects to help overcome barriers to effective innovation. BRIDGE strives to provide unified and clear advice in this regard, in order to reinforce the messages and optimize their influence on policymakers.

Use case-based approaches to business and functional processes, the Common Information Model (CIM) for standardised data exchange (semantics and syntactics), and the Harmonized Electricity Market Role Model (HEMRM) for clearly defining roles and responsibilities are all building blocks for the definition of a common European reference architecture.

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

The BRDIGE RA is the outcome of mapping the results to the interoperability layers of the smart grid architectural model (SGAM). While features relevant to the electrical domain are emphasized on the left side of the picture, the focus of this study and reference architecture is on the centre section - the cross-sector domain. This is where the most interoperability efforts are required.

Data interchange is what makes cross-sector interoperability possible. In the information and communication levels, the data aspect is extremely natural. However, it may and should be the case in the business, function, and component layers as well. As a result, cross-sector aspects should be the focus of collaborative and well-coordinated initiatives.

The figure focuses on features that were cited in BRIDGE initiatives or were somehow retrieved as significant to the cross-sector area. Clearly, the list of factors listed isn't exhaustive.

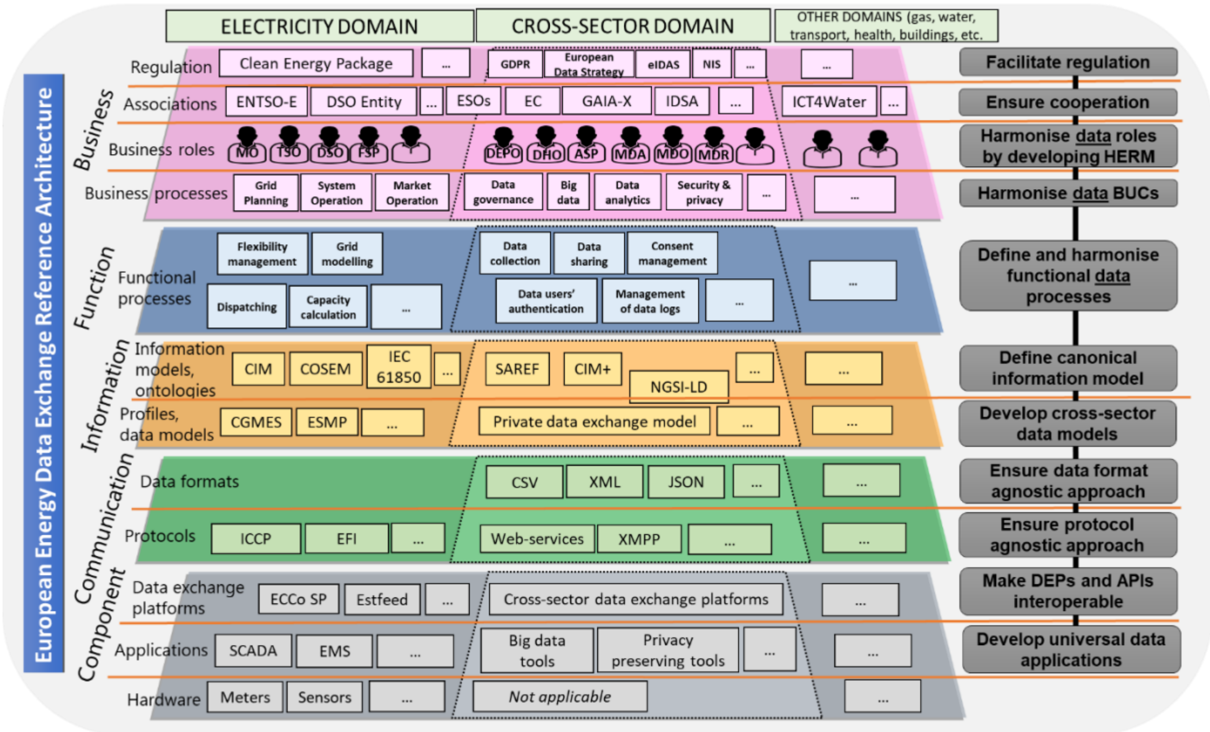


Figure 2 BRIGE Reference Architecture Framework

All of the pieces are displayed using SGAM layers, which are further divided into sub-layers based on the data interchange requirements. These layers are described as follows:

The Business Layer, which reflects the business perspective on smart grid information sharing, is divided into four parts:

- Regulation: it contains the different regulations in different domains.
- Associations: ENTSO
- Business Roles
- Business Processes

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

The Function Layer includes the functional processes such as grid modelling and flexibility management, which specifies functions and services as well as their interactions from an architectural standpoint.

The Information Layer represents the data that is shared and utilized by functions, services, and components. It comprises information objects as well as the underlying canonical data models, which describe the common semantics for functions and services in order to facilitate interoperable data exchange across communication channels. It is divided into the following parts:

- Information Models, ontologies
- Profiles, data models.

The Communication Layer is divided into two parts, each of which describes protocols and techniques for the interoperable flow of information between components in the context of the underlying use case, function, or service, and related information objects or data models:

- Data Formats
- Protocols

The Component Layer, which includes system actors, applications, power system equipment, protection and tele-control devices, network infrastructure (wired / wireless communication connections, routers, switches, servers), and any kind of computers, is split into:

- Data exchange platforms
- Applications
- Hardware.

On the right side of the architecture in Figure 2, recommendations are offered for each layer and sub-layer. Because data source integration is such an important part of the utilities, it's vital to develop interdependencies and data flow flows among them. Furthermore, the architecture incorporates different industries (e.g., gas, heating, water, etc.). From the architecture models covered in this section, the BRIDGE RA presents the most comprehensive view of use cases in the electrical sector, as well as cross-sectorial use cases. It is based on the widely used Smart Grid Architecture Model (SGAM).

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

3 MATRYCS Stakeholders and Requirements Analysis

The requirements and design principle has already described in the D2.3, which contains also the MATRYCS use cases from MATRYCS Deliverable D2.1 together with the specifications and requirements from D2.2. In this final Deliverable D2.4, the updated version of the user stories, usage scenarios and use cases are descripted.

These are identified considering: i) already existing works on the design of Reference Architectures (presented in Section 2), since they contain aspects that have general relevance, regardless of the specific field or domain of application; ii) the MATRYCS use cases (presented in the MATRYCS Deliverable D2.1) together with the associated specifications and requirements (presented in the MATRYCS Deliverable D2.2), since they include needs or aspects specific to the building sector, being relevant for the discussion of building data spaces and for the definition of a Reference Architecture adapted to the building domain.

Keeping this twofold view in mind, in the following, this Section will first present the roles of the different stakeholders involved in the process of value creation for the building data, with the goal of identifying specific technical or business needs that should be considered in the Reference Architecture definition. Then, the existing functional and non-functional requirements will be analysed, also based on the considerations derived from the use cases proposed in the MATRYCS Large Scale Pilots. Finally, some more general design principles and recommendations will be discussed, which are at the basis of the definition of the MATRYCS Reference Architecture.

3.1 MATRYCS User Stories, Usage Scenarios and Use Cases

This section presents the updated version of the User Stories, Usage Scenarios and Use Cases, as the main outcome of T2.2 - "User Stories and Requirements Analysis". For this, a brief summary of the process followed is presented in section 3.1.1 (the in-depth description can be found in D2.1 - "State of the art analysis and Big Data Value Chain"). Then, the final list of use cases is shown in section 3.1.2, and, finally, the results are quantified in section 3.1.3.

To complement this section, ANNEX I presents final list of use cases is related to their corresponding usage scenarios and user stories; and also, the discarded usage scenarios and user stories are compiled.

3.1.1 Description of the process

The process for usage scenarios, use cases and user requirements was defined at the early stages of MATRYCS project. The objective is to produce the Service Requirements but without losing the user perspective or needs, that is, starting from the LSP language and goals and then be able to match it with the MATRYCS services, and finally produce the Requirements.

The process began by defining a **Test Card** by each LSP, with the aim of making the Pilots consider and reflect on their needs and objectives in the Project (and reformulate something of what was stated in the proposal, if necessary). The Test Card was a simple card per LSP including the LSP name, their

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

main goals, several metrics that can help to determine if the goals are achieved, and the definition of Target Groups and related Personas (main stakeholders benefited by the Pilot development).

Then, the **Personas** identified in the Test Card of the LSP were deeply and individually analysed, describing for each one their user context, pain points (that service could solve), goals, usefulness of the service ('how the service fit into their life'), and attitudes and sentiments (the emotions that service will create).

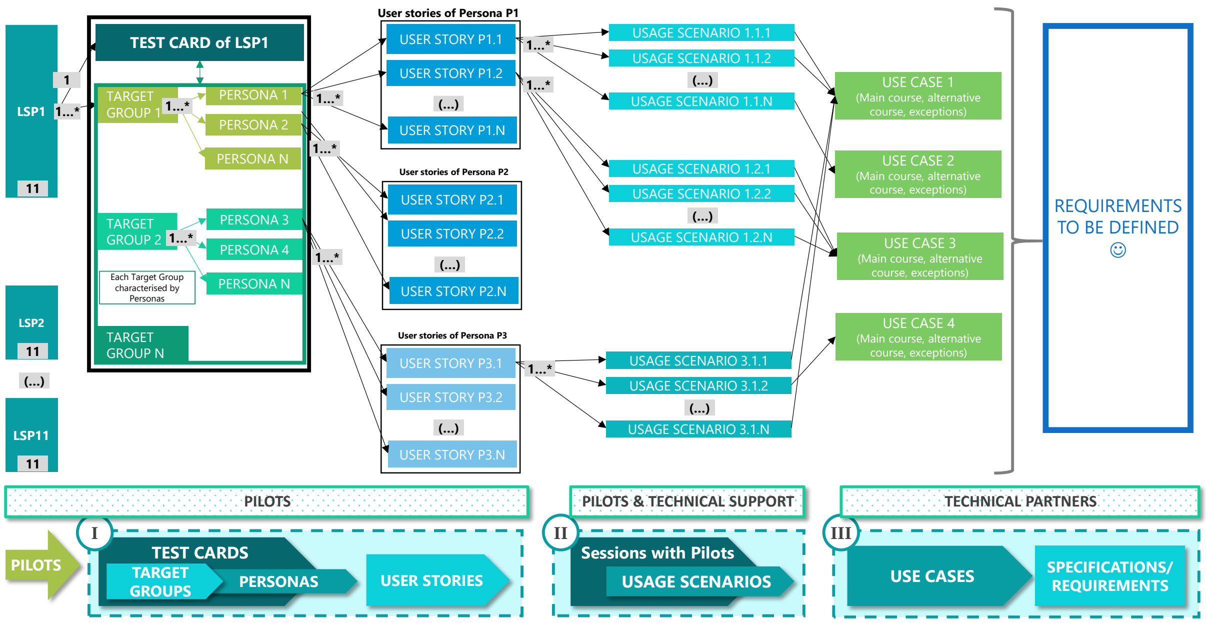
This Personas detailed analysis help then the creation of the **User Stories**, which are formulated as a simplification of their main goals. The User Story is a method used to capture a description of a service feature from the user's perspective. They are written in a natural language and placed at the center of the user experience. A User Story has the following structure:

- As a [ROLE] (**Who** it is for? → **Persona** from which this User Story is originated), I want to [ACTION] (**What** it expects from the system? → extracted from the main **pains and goals** from the Persona) so that I can [BENEFIT] (**Why** it is important? → extracted from the **usefulness** of the service for the Persona, specific **benefits** from the service).

Considering that for each Persona, multiple goals could have been formulated, there were usually several User Stories for each Persona.

After the User Stories, the **Usage Scenarios** were created. They are detailed versions of the User Stories, covering the involved Persona, specific pre-conditions and post-conditions, as well as the several steps or actions needed to the accomplishment of the Usage Scenario. Then, a review and relation were made between the steps of the Usage Scenarios and the MATRYCS services, to match what of them can actually be covered by them.

The **Use Cases** were finally generated from all previous inputs. The Usage Scenarios were clustered or grouped and the same or similar needs were grouped into types. The same type Usage Scenarios led to the final Use Case, which is already closer to the statement of Service requirements.



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

Figure 3: The complete picture of use cases and user requirements process generation and matching

3.1.2 Final use cases, user stories and usage scenarios

This section presents the final list of use cases, which have been updated from those presented in D2.1 "State of the art analysis and Big Data Value Chain". It should be mentioned that in this document only basic information of the use cases is presented (scope, objective and comments). The complete versions of these use cases following the template presented in D2.1 (section 5.1.3) will be presented in D6.14 "Report on Pilot's Integration and Validation".

As an introduction to the final list of use cases, is worth to highlight the final services that are to be applied in each LSP. Some minor modifications have been performed from the version presented in D2.1, mainly due to the problems encountered with the data available which have constrained the application of services in particular LSPs.

Table 1: MATRYCS services / pilots' matrix (updated)

		MATRYCS Large Scale Pilots										
		LSP1	LSP2	LSP3	LSP4	LSP5	LSP6	LSP7	LSP8	LSP9	LSP10	LSP11
MATRCYS SERVICES	s0.1		X				X		X	X		
	s0.2								X		X	
	s1.1	X			X	X	X					
	s1.2	X			X	X						
	s1.3		X				X		X	X		
	s1.4	X										
	s1.5				X		X					
	s2.1					X						
	s2.2		X									
	s3.1							X				
	s3.2									X		
	s3.3											X
	s4.1			X			X					
	s4.2		X						X		X	

Based on the above table, the following Table 2 summarises the use cases, relates them to the services that are involved and quantifies the number of scenarios contained in each use case as well. It must be highlighted that these "scenarios" are a simplified version of the usage scenarios developed in the process.

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

Table 2: MATRYCS use cases, usage scenarios, user stories and services applied

#	UC ID	Use case name	Scenarios	Nº Usage scenarios	Nº User stories	LSP	S0.1	S0.2	S1.1	S1.2	S1.3	S1.4	S1.5	S2.1	S2.2	S3.1	S3.2	S3.3	S4.1	S4.2	Use cases /LSP
1	UC01_01	Action plans for preventive maintenance	2	4	4	LSP1				X		X									2
2	UC01_02	Sustainability measures for performance improvements	2	7	7				X	X											
3	UC02_01	Building evaluation of ECM / Building refurbishment	1	2	2	LSP2									X					X	2+2
4	UC02_02	Visualization of indicators in the Digital Twin	1	2	2						X										
5	UC03_01	M&V plan creation	1	3	3	LSP3													X		2
6	UC03_02	Savings report generation	2	3	3														X		
7	UC04_01	Optimization of electrical distribution system operation	1	3	3	LSP4			X				X								2
8	UC04_02	Sustainability measures for performance improvements	1	3	2				X	X											
9	UC05_01	Predictions about energy	1	3	3	LSP5			X												4
10	UC05_02	Identification of PV performance issues	1	2	2				X												
11	UC05_03	Optimization of self-consumption of PV system	1	4	4				X	X											
12	UC05_04	Creation of technology catalogues	1	1	1									X							
13	UC06_01	M&V plan creation	1	1	1	LSP6													X		4+2
14	UC06_02	Savings report generation	2	1	1														X		
15	UC06_03	Operation of DH Network	1	7	7				X				X								

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

16	UC06_04	Designing an extension of the district heating network	2	5	5				X		X		X								
17	UC07_01	Policy-makers decision-making support	1	7	7	LSP7										X					1
18	UC08_01	Geo_municipality service	1	1	1	LSP8		X													3+2
19	UC08_02	Residential building energy consumption estimation	1	2	2						X										
20	UC08_03	Risking analysis in energy efficiency	1	2	2															X	
21	UC09_01	EPCs data error checker and validator	1	2	2	LSP9											X				3+2
22	UC09_02	Visualization of estimated EPC	2	5	5						X						X				
23	UC09_03	Visualization of the energy building improvements from the EPCs	1	3	3												X				
24	UC10_01	ECM and investment	1	3	3	LSP10		X													2
25	UC10_02	Risk assessment of funding in energy efficiency investments	1	3	3															X	
26	UC11_01	Portfolio optimization	1	2	2	LSP11												X			1
27	UC00_01	Digital Twin creation	3	N/A	N/A	LSP2 LSP6 LSP8 LSP9	X														(2)
28	UC00_02	Services integration in Digital Twin	3	N/A	N/A		X		(X)		(X)		(X)				(X)				

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

The following table presents the latest version of the use cases, by defining their scope, objectives and the scenarios considered in each of them.

Table 3: MATRYCS use cases identified so far

#	Use Case ID	Name	Related LSPs	Involved services	Actors
1	**UC01_01**	***Action plans for preventive maintenance***	*LSP1*	*s1.2, s1.4*	*Facility Managers*
Scope	*Definition of energy efficiency measurements and to provide support for preventive and predictive maintenance of the systems at building level.*				
Objective	*Define suitable solutions for providing decision support to improve sustainable aspects and improve thermal comfort. Additionally, to develop technical functionalities for building management, fault detection and diagnostics.*				
Scenario(s)	*1*	*<u>Name:</u> List of action plans for preventive maintenance* *<u>Corresponding USC: LSP1:</u> USC 2.2.1, USC 3.1.1, USC 3.2.1* *<u>Actors</u>: Facility Manager; Cost-conscious facility manager; Sustainability supporting facility manager*			
	2	*<u>Name</u>: Support for preventive and predictive maintenance of the systems* *<u>Corresponding USC: LSP1</u>: USC 1.2.1, USC 2.2.1, USC 3.1.1, USC 3.2.1* *<u>Actors</u>: Facility Manager; Cost-conscious facility manager; Sustainability supporting facility manager*			
2	**UC01_02**	***Sustainability measures for performance improvements***	*LSP1, LSP4*	*s1.1, s1.2*	*Facility managers, Sustainability supporting owner, Informed owner of shops, Sensitive office employee*
Scope	*Identification of possible action plans and selection of optimal solution to provide energy savings considering sustainability aspects and thermal comfort*				
Objective	*Evaluate and report performance measurements*				
Scenario(s)	*1*	*<u>Name:</u> Actions plans for performance improvements* *<u>Corresponding USC</u>: <u>LSP1</u>: USC 1.1.1, USC 5.1.1, USC 5.3.1, USC 2.1.1* *<u>Actors</u>: Facility manager; Sustainability supporting owner; Informed owner of shops*			
	2	*<u>Name:</u> Actions plans for performance improvements including thermal comfort aspects* *<u>Corresponding USC</u>: <u>LSP1</u>: USC 1.4.1, USC 4.1.1, USC 5.2.1. <u>LSP4</u>: USC 1.1.1, USC 1.1.2* *<u>Actors</u>: Informed facility manager; Sustainability supporting owner; Informed owner of shops; Sensitive office employee*			
3	**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

Scope	*Identification of list of possible renovation actions to improve energy efficiency at building level*				
Objective	*Definition of priorities for the possible refurbishment actions to enhance the energy efficiency of buildings and provision of associated economic information*				
Scenario(s)	*1*	*Name: ECM evaluation in building Corresponding USC: LSP2: USC 2.1.1, 3.1.1 Actors: Management staff of buildings, designers*			
4	**UC02_02**	***Visualization of indicators in the Digital Twin***	*LSP2*	*s1.3*	*Management staff of buildings, Workers of a municipality*
Scope	*Selection of relevant indicators to be displayed in the corresponding digital twin*				
Objective	*Define the best way to report on the chosen indicators in the digital twin environment: How to access these indicators, how to display them (point values, evolution) and possible operations to perform with them (report generation, export). KPIs presented in this way can provide a quick overview of the performance of the building.*				
Scenario(s)	*1*	*Name: Key Performance Indicators visualization in Digital Twin framework Corresponding USC: LSP2: USC 1.1.1, 2.2.1 Actors: Management staff of educational buildings, Workers of a municipality*			
5	**UC03_01**	***M&V plan creation***	*LSP3, LSP6*	*s4.1*	*Advanced ESCO Informed owner*
Scope	*Selection of the best mathematical model to predict energy savings.*				
Objective	*Define M&V plan based on IPMVP (Option C). Selection of the best mathematical model to predict energy savings.*				
Scenario(s)	*1*	*Name: M&V plan creation Corresponding USC: LSP3: USC 1.1.1, USC 1.4.1, USC 2.3.1; LSP6: USC 6.1.1 Actors: Advanced ESCO, Informed owner*			
6	**UC03_02**	***Savings report generation***	*LSP3, LSP6*	*s4.1*	*Advanced ESCO, Informed owner*
Scope	*Verification of the savings achieved after the implementation of Energy Conservation Measures (ECMs) to a building/district through a recognized measurement and verification protocol (IPMVP).*				
Objective	*Verify and report the savings achieved after the implementation of ECMs.*				
Scenario(s)	*1*	*Name: Basic savings report generation Corresponding USC: LSP3: USC 1.4.1, USC 2.3.1; LSP6: USC 6.1.1 Actors: Advanced ESCO, Informed owner*			
	2	*Name: Savings report generation Corresponding USC: LSP3: USC 1.1.1 Actors: Advanced ESCO*			
7	**UC04_01**	***Optimization of electrical distribution system***	*LSP4*	*s1.1, s1.5*	*Small Scale DSO Head*

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

		operation			
Scope	*To monitor and optimise the electrical distribution system operation*				
Objective	*To understand and predict accurately the behaviour of the system operation so as to be optimised and avoid volatilities*				
Scenario(s)	*1*	*Name: Optimization of electrical distribution system operation* *Corresponding USC: LSP4: USC 7.1.1, USC 7.2.1, USC 7.3.1* *Actors: Small Scale DSO Head*			
8	**UC04_02**	***Sustainability measures for performance improvements***	*LSP1, LSP4*	*s1.1, s1.2*	*Facility managers, Sustainability supporting owner, Informed owner of shops, Sensitive office employee*
Scope	*Identification of possible action plans and selection of optimal solution to provide energy savings considering sustainability aspects and thermal comfort*				
Objective	*Evaluate and report performance measurements*				
Scenario(s)	*1*	*Name: Sustainability measures for performance improvements including thermal comfort aspects* *Corresponding USC: LSP4: USC 1.1.1, USC 1.1.2, USC 1.2.1* *Actors: Informed facility manager; Sustainability supporting owner; Informed owner of shops; Sensitive office employee*			
9	**UC05_01**	***Predictions about energy***	*LSP5*	*s1.1*	*EV owner; eco-friendly SME*
Scope	*Implementation of several energy prediction applications, depending on the input data.*				
Objective	*The final benefit is a high-quality prediction tool, based on latest techniques and applications to calculate carbon footprint.*				
Scenario(s)	*1*	*Name: Predictions about energy* *Corresponding USC: LSP5: USC 2.1.1, USC 2.2.2, USC 5.1.1* *Actors: EV owner; eco-friendly SME*			
10	**UC05_02**	***Identification of PV performance issues***	*LSP5*	*s1.1*	*Prosumer*
Scope	*To identify when a PV system is not performing properly*				
Objective	*Identification of possible maintenance actions in a PV system and its overall efficiency*				
Scenario(s)	*1*	*Name: Identification PV performance issues* *Corresponding USC: LSP5: USC 1.1.1, USC 1.1.2* *Actors: Prosumer*			
11	**UC05_03**	***Optimization of self-consumption of PV system***	*LSP5*	*s1.1, s1.2*	*Prosumer*
Scope	*To maximize the value of an investment of a PV system*				

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

Objective	*Increase of self-consumption and reduction of surplus production of the system*				
Scenario(s)	*1*	*Name: Optimization of self-consumption with PV system* *Corresponding USC: LSP5: USC 1.2.1, USC 1.2.2, USC 1.2.3, USC3.1.1* *Actors: Prosumer*			
12	***UC05_04***	***Creation of technology catalogues***	*LSP5*	*S2.1*	*Environmentally conscious citizens*
Scope	*Generation of standardised technologies catalogues to support the design of optimal solutions at building or district level of a community.*				
Objective	*The outcome will be a tool that provides list of potential technologies that can be applied based on building/ districts and area specific characteristics. This will allow members of a community to make better decisions, when designing renovation road maps at the local level (specific blocks of buildings/ district).*				
Scenario(s)	*1*	*Name: Catalogue creation* *Corresponding USC: LSP5: USC 3.2.1* *Actors: Environmentally conscious citizens*			
13	**UC06_01**	***M&V plan creation***	*LSP3, LSP6*	*s4.1*	*Advanced ESCO Informed owner*
Scope	*Selection of the best mathematical model to predict energy savings.*				
Objective	*Define M&V plan based on IPMVP (Option C). Selection of the best mathematical model to predict energy savings.*				
Scenario(s)	*1*	*Name: M&V plan creation* *Corresponding USC: LSP3: USC 1.1.1, USC 1.4.1; LSP6: USC 6.1.1* *Actors: Advanced ESCO, Informed owner*			
14	**UC06_02**	***Savings report generation***	*LSP3, LSP6*	*s4.1*	*Advanced ESCO, Informed owner*
Scope	*Verification of the savings achieved after the implementation of Energy Conservation Measures (ECMs) to a building/district through a recognized measurement and verification protocol (IPMVP).*				
Objective	*Verify and report the savings achieved after the implementation of ECMs.*				
Scenario(s)	*1*	*Name: Basic savings report generation* *Corresponding USC: LSP3: USC 1.4.1; LSP6: USC 6.1.1* *Actors: Advanced ESCO, Informed owner*			
	2	*Name: Savings report generation* *Corresponding USC: LSP3: USC 1.1.1* *Actors: Advanced ESCO*			
15	**UC06_03**	***Operation of DH Network***	*LSP6*	*s1.1, s1.5*	*Sustainability supporter DHN manager, Cost-conscious facility manager, Efficiency-oriented DHN manager*

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

Scope	*To effectively understand the operation of the DH Network*				
Objective	*To predict accurately the behaviour of the facility, its response to possible changes and increase the knowledge about the operation of the facility.*				
Scenario(s)	*1*	*Name: Operation of DH Network* *Corresponding USC: LSP6: USC 1.1.1, USC 1.2.1, USC 1.3.1, USC 2.1.1, USC 2.2.1, USC 2.3.1, USC 4.3.1* *Actors: Sustainability supporter DHN manager, Cost-conscious facility manager, Efficiency-oriented DHN manager*			
16	**UC06_04**	***Designing an extension of the district heating network***	*LSP6*	*s1.1, s1.3, s1.5*	*Profit-oriented DHN manager, Efficiency-oriented DHN manager, Informed investor*
Scope	*To understand the impact in the operation of the DH Network of adding new buildings to the network.*				
Objective	*To predict accurately the behaviour of the new buildings to be integrated in the existing district heating network; and to analyse the whole future facility, its response to possible changes and increase the knowledge about the operation of the facility.*				
Scenario(s)	*1*	*Name: Calculating the behaviour of new buildings in the DHN (s1.3)* *Corresponding USC: LSP6: USC 3.1.1, 3.2.1, 4.1.1, 4.2.1, 5.3.1* *Actors: Profit-oriented DHN manager, Efficiency-oriented DHN manager, Informed investor*			
	2	*Name: Predicting the behaviour and analysing of operation the future DHN (s1.1 + s1.5)* *Corresponding USC: LSP6: USC 3.1.1, 3.2.1, 4.1.1, 4.2.1, 5.3.1* *Actors: Profit-oriented DHN manager, Efficiency-oriented DHN manager, Informed investor*			
17	**UC07_01**	***Policy-makers decision-making support***	*LSP7*	*s3.1*	*Urban Planning Professional, Local department director*
Scope	*Support policy-makers regarding climate change initiatives and action plans.*				
Objective	*The outcome will be a tool that enables policy-makers to gain valuable insights and make better decisions when they plan and design action plans to tackle climate change. This tool will be enabled by visualizations, comparisons and aggregation based on existing data and possibly data that the user will provide.*				
Scenario(s)	*1*	*Name: Policy-makers decision-making support (and assessment)* *Corresponding USC: LSP7: USC 1.1.1, USC 1.2.1, USC 2.1.1, USC 2.2.1, USC 3.2.1, USC 4.1.1, USC 4.2.1* *Actors: Urban Planning Professional, Local department director*			
18	**UC08_01**	***Geo_municipality service***	*LSP8*	*s0.2*	*Real state owner, Workers of the municipality and other municipalities r*

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

Scope	*Show information regarding the performance of public building (School building available) based on energy consumption data provided by the municipality.*				
Objective	*1) Visualize specific information about energy performance of public buildings* *2) Workers use it to compare the performance of different buildings.* *3) The KPIs can provide information about the possibility to refurbish the building.*				
Scenario(s)	*1*	*Name: Information on energy performance of municipal public buildings* *Corresponding USC: LSP8: USC 1.1.1, USC 4.2.1* *Actors: Real state owner, Workers of the municipality and other municipalities*			
19	**UC08_02**	***Residential building energy consumption estimation service***	*LSP8*	*S1.3*	*Real state owner, Workers of the municipality and other municipalities*
Scope	*To show information regarding the consumption of residential buildings based on calculation using public and open data*				
Objective	*To calculate using EPBD-based algorithms estimation for consumption of residential buildings using public information. To show this information geo-located in a map*				
Scenario(s)	*1*	*Name: Residential building energy consumption estimation* *Corresponding USC: LSP8: USC 1.1.1, USC 4.1.1* *Actors: Real state owner, Workers of the municipality and other municipalities*			
20	**UC08_03**	***Risking analysis in energy efficiency***	*LSP8*	*S4.2*	*Real state owner, Building administrators and managers*
Scope	*Give the possibility to calculate the possible risk in investment in energy efficiency*				
Objective	*The outcome will be a tool by means the user can visualize the possible risk on investment in energy efficiency using the engine developed in the EEnvest project*				
Scenario(s)		*Name: De-risking in energy efficiency* *Corresponding USC: LSP8: USC 1.1.1, USC 2.1.1* *Actors: Real state owner, Building administrators and managers*			
21	**UC09_01**	***EPCs data error checker and validator***	*LSP9*	*s3.2*	*Energy Performance Certification (EPC) issuers*
Scope	*Support EPC issuers in detecting errors in the data entry to EPC softwares.*				
Objective	*The outcome will be a relevant tool that enables EPC issuers to offer trustful calculations on energy consumption and saving and also Regional Administration civil servant (EPC register) to verify registered EPCs detecting potential errors increasing the quality of EPC's Regional Administration data sets*				
Scenario(s)	*1*	*Name: EPCs data error checker and validator* *Corresponding USC: LSP9: USC 1.1.1, USC 2.1.1* *Actors: Energy Performance Certification (EPC) issuers*			
22	**UC09_02**	***Visualization of***	*LSP9*	*s3.2, s1.3*	*Regional Energy*

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

		estimated EPC			*Efficiency planner*
Scope	*The scope is the generation of estimated EPCs based in different approaches and providing this information in different levels of detail to different stakeholders*				
Objective	*The outcome will be a relevant tool that will allow the energy efficiency planner to obtain valuable insights from the registered buildings EPCs, for making better decisions when designing action plans and also will provide information of estimated EPCs to different actors: Energy Performance Certification (EPC) issuers, buildings buyers and sellers, landlords and tenants.*				
Scenario(s)	*1*	*Name: Visualization of estimated EPC Corresponding USC: LSP9: USC 1.2.1, 4.1.1 Actors: Regional Energy Efficiency planner*			
	2	*Name: Visualization of estimated EPC advanced Corresponding USC: LSP9: USC 2.3.1, 3.1.1, 3.2.1 Actors: Regional Energy Efficiency planner*			
23	**UC09_03**	***Visualization of the energy building improvements from the EPCs***	*LSP9*	*S3.2*	*Regional Administration civil servant, citizens, tenants and Owners*
Scope	*To offer information about potential energy building improvements to be applied in buildings taking into account those measures indicated in the EPCs.*				
Objective	*To gather all the energy building improvements available in the EPCs. To show this information in a visual and structured way in one useful way for the end-users.*				
Scenario(s)	*1*	*Name: Visualization of estimated EPC Corresponding USC: LSP9: USC 2.2.1, USC 3.3.1, USC 4.2.1 Actors: Regional Administration civil servant, citizens, tenants and Owners*			
24	**UC10_01**	***ECM and investment***	*LSP10*	*S0.2*	*Informed policy maker*
Scope	*Make decisions quickly and without spending a lot of time analysing information, knowing which the best refurbishment solution for each building is.*				
Objective	*Identify quickly based on the data provided what action can be taken to achieve certain energy savings.*				
Scenario(s)	*1*	*Name: ECM and investment Corresponding USC: LSP10: USC 5.1.1, USC 5.2.1, USC 5.3.1 Actors: Informed policy maker*			
25	**UC10_02**	***Risk assessment of funding in energy efficiency investments***	*LSP10*	*S4.2*	*Informed policy maker*
Scope	*To assess risk evaluation of funding in energy efficiency investments.*				
Objective	*To define reliable investments, to identify opportunities and make comparisons between different energy efficiency actions and to have a better view of funding direction and proposals.*				
Scenario(s)	*1*	*Name: Financing risk assessment Corresponding USC: LSP10: USC 5.1.1, USC 5.2.1, USC 5.3.1*			

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

		Actors: Informed policy maker			
26	**UC11_01**	***Portfolio optimisation***	*LSP11*	*s3.3*	*National authority - policy maker*
Scope	*Support policy making, by providing optimal solutions under a ranging number of different criteria.*				
Objective	*The outcome of the service will be a graphical representation of optimal portfolios, which features a different pair of cost-risk values. For each selected action, the number the number of interventions is extracted from the optimisation process. For each intervention of the selected actions, the year of implementation is also decided from the optimisation process.*				
Scenario(s)	*1*	*Name:* Portfolio optimisation *Corresponding USC:* LSP11: USC 1.1.1, USC 2.1.1 *Actors:* National authority policy maker			
27	**UC00_01**	***Digital Twin creation***	*LSP2, 6, 8, 9*	*s0.1*	*DT creator; Service developer*
Scope	*Analytics services and Digital Twin definition at Building, District and City/ Region scale. Compile the relevant information to be used at three different scales to support the implementation of energy services.*				
Objective	*The objective is to support the deployment of other services by providing a digital twin, which integrates all relevant information at each scale. This will allow to automatically generate models based on other data sources if they do not exist. To this purpose, s0.1 Digital Twin is involved.*				
Scenario(s)	*1*	*Name:* DT creation at building scale *Corresponding USC:* N/A *Actors:* DT creator, Service developers			
	2	*Name:* DT creation at district scale *Corresponding USC:* N/A *Actors:* DT creator, Service developers			
	3	*Name:* DT creation at city/ regional scale *Corresponding USC:* N/A *Actors:* DT creator, Service developers			
28	**UC00_02**	***Services integration in Digital Twin***	*LSP2, 6, 8, 9*	*s0.1*	*DT creator; Service developer*
Scope	*Analytics services integration in Digital Twin*				
Objective	*Implement in Digital Twins analytics services to feed these services with data and enrich the digital twin with the outputs generated in the services. The objective is to integrate analytics services of the same type in the Digital Twins answering the needs specified in the pilots and captured in the user stories and the specifications.* *Prepare/adapt the Digital Building Twins to the required inputs and outputs, the required predictions and behavioural analysis, as well as the features to be applied for the ML and DL definitions.*				
Scenario(s)	*1*	*Name:* KPIs calculation			

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

		Corresponding USC: N/A *Actors: DT creator; Service developer*
	2	*Name: District network support* *Corresponding USC: N/A* *Actors: DT creator; Service developer*
	3	*Name: GIS support* *Corresponding USC: N/A* *Actors: 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

3.1.3 Results and statistics

As a summary, the following Table 4 presents the statistics with respect to the scenarios, target groups, personas, user stories and final usage scenarios considered. Also, the discarded usage scenarios are presented. The reason for discarding some of them, in most cases was the more precise alignment of the LSP to its specific goals and be able to maintain a balance with the amount of effort depicted in the Grant Agreement. As it can be seen, the Scientific Objective 1 "Design of at least 11 different usage scenarios for the building sector" is clearly achieved, since there are a total of 32 scenarios included in the use cases defined in MATRYCS.

Table 4: MATRYCS user stories, usage scenarios, use cases statistics

LSP	Scenarios	Nº Target groups	Nº Personas	Nº User stories	Nº Final usage scenarios	Nº discarded usage scenarios
LSP1	4	2	6	11	11	9
LSP2	2	2	3	4	4	0
LSP3	3	1	2	3	3	16
LSP4	2	2	2	5	6	4
LSP5	4	2	4	7	10	6
LSP6	6	3	6	12	13	5
LSP7	1	2	4	7	7	5
LSP8	3	3	3	3	4	5
LSP9	4	3	4	10	10	2
LSP10	2	1	1	3	3	15
LSP11	1	2	2	2	2	1
TOTAL	32	23	37	67	73	68

3.2 MATRYCS requirements and specifications

This Section recalls the MATRYCS requirements and specifications extracted from the MATRYCS LSP use cases, as reported in the MATRYCS Deliverable D2.2 [4]. In D2.2, an in-depth analysis of the use cases was performed to understand the associated functional and non-functional requirements. The same process was followed after the 1st version of the MATRYCS Reference Architecture (D2.3) and the 1st technology release of the MATRYCS layers, modules, components, services and MATRYCS Analytics Toolbox (D3.1 [6], D4.1 [7], D4.4 [8] and D5.1 [9]). This multi-level process included information from updated Use Cases (Section 3.2) and consequently in conjunction with LSP information, requirements and services to be deployed. The updated and final requirements' list were clustered according to their typology (functional and non-functional) and category (analytics, collection, compatibility, integration,

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

interaction, notification, reporting, visualisation, modularity, availability, efficiency, interoperability, reliability, security, usability). The "saving" category is not finally included as a required functionality of the MATRYCS Reference Architecture as the MATRYCS modules and services development was done so as the end-user to receive generated data but not stored them back in MATRYCS data base. The update list of requirements is still following the macro-requirements from other Architecture's proposal as analysed in D2.3.

The updated list of the User Requirements was reduced by 14 IDs. The final version of MATRYCS reference architecture was based on 92 user requirements (65 functional and 21 non-functional). The deleted IDs were resulted from deleted Use Cases or merges from similar IDs. Finally, after the finalisation of MATRYCS technical layers and especially on the visibility and interaction aspects, several IDs were rejected as their required functionality is not provided to the end-users.

It should be noted that the User Requirement elicitation includes all Use Cases mapped at least at one User Requirement, while general User Requirements are used and refer to all Use Cases and services, Table below summarises the User Requirements characteristics and their type while Anexx II provide the full list of the final User Requirements.

Table 5: List of MATRYCS functional and non-functional requirement categories

Requirement Category	#URs	Description	Type
Analytics	24	Need of advanced data analytics functionalities aimed at enhancing the information associated to the input data	Functional
Collection	11	Need of dedicated pre-processing routines aimed at cleaning, filtering or integrating the raw data provided by the physical system or by end-users	Functional
Compatibility	5	Need of having interconnectivity and interoperability among specific technology components in the ecosystem	Functional
Integration	2	Need of having syntactic and semantic interoperability	Functional
Interaction	4	Need for end-users to freely query, visualize and interact with the ecosystem data they have access to	Functional
Notification	2	Need of specific functionalities aimed at notifying the end-used in case of occurrence of specific events	Functional

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

Reporting	13	Need of specific functionalities able to present in a clear and effective way the information extracted from sets of data	Functional
Visualization	4	Need of specific functionalities aimed at presenting data or service results in an effective and attractive form	Functional
Modularity	1	Feature for which the components and modules of the overall ecosystem are able to work independently	Non-functional
Efficiency	4	Feature for which the overall system or its components are able to perform complex operations in a relatively short time	Non-functional
Interoperability	4	Feature for which the system is able to successfully collect input data coming from the physical world or from other systems	Non-functional
Reliability	5	Feature for which the system is still able to provide its fundamental functionalities also in presence of unexpected errors or failures	Non-functional
Security	4	Feature for which the system is capable to prevent any unauthorized operation towards data and system components	Non-functional
Usability	9	Feature for which end-user can easily use the functionalities provided by the system	Non-functional

3.2.1 Security

The security aspects of the MATRYCS Toolbox were first introduced in deliverable D2.2 - "MATRYCS Technical & Security Specifications" and then updated in deliverable D2.3 - "MATRYCS Reference Architecture for Buildings Data v1.0". These aspects are being monitored and will continuously assessed be and verified throughout the project. The development of the technical components and services that will be available through MATRYCS, necessitate the conformation to the security aspects according to the specifications of the project. In this section a summary of the updates that have emerged is provided.

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

- **Data Confidentiality:** The concept of confidentiality indicates that each dataset provided during the data sharing procedure, will be accessed by the respective authorized data consumers. This principle is strictly followed throughout data sharing activities, as each LSP share its data through a dedicated folder in the project's MS Teams repository. Data access limitations have been guaranteed, as the only partners that can access and modify these datasets are the technical partners related to the respective pilot (data controller role, as described in deliverable D2.2 - "MATRYCS Technical & Security Specifications). Moreover, all the software components that have been developed in WPs 3, 4 and 5 are implemented and maintained in Git environments, where access is guaranteed only to MATRYCS technical partners.

  In the context of the MATRYCS Toolbox, data confidentiality is secured via an enhanced Security Layer, implemented in the context of the MATRYCS End-to-End Security Framework. This module prevents unauthorized users from accessing MATRYCS datasets of other pilot partners, assuring that pilots' data will not be accessible through the services from non-authorized users. In this respect, only the specified users will have access to the services developed through unique passwords under a role-based approach.

- **Data Integrity:** Data integrity is monitored through activities of WP3. Raw data are pre-processed and harmonized before being transferred to the staging area. This procedure reserves that data will be uploaded in the MATRYCS data in a consistent and harmonized way, conforming to certain principles, that are carefully designed for each pilot's data, and in collaboration with the respective service(s) needs.

- **Data Availability:** Data are guaranteed to be accessible by authorized users when demanded. At this stage of the project, this is implemented in three levels including: (a) raw, unprocessed data in .csv or .xlsx format stored in the MS Teams Repository, (b) harmonized pre-processed data available in the staging area of the project and (c) final format of the data available and accessible through the MATRYCS Database. The final procedures for data availability, will be clarified and reported under WP3 activities.

  The first level (Teams Repository as of now) enables each pilot to easily share samples of data in a quick way and allow the respective technical contact partners and service providers have a clear picture of the format of the data in the initial phase of the MATRYCS project.

  The second level (staging area) includes the harmonized format of data after pre-processing has been finalized. This layer enables data to be easily accessible and available between technical partners during the project and before they are finally integrated into the MATRYCS database.

  Finally, the third level (MATRYCS databases) is the point of interconnection between the technical WPs 3, 4 and the MATRYCS Analytics WP5. The MATRYCS services will query data through the database to provide visualizations and statistics, and to generally connect data to users. Thus, data availability in this layer is very important, being closely related to the development and smooth operation of the MATRYCS services.

- **Data Reliability:** As a cloud-based system. MATRYCS platform will innately have the capability of avoiding single points of failure, providing fundamental functionalities in case of unexpected anomalies and errors. This is also clearly stated and followed under the technical requirements of the MATRYCS project.

3.2.2 Privacy

Privacy requirements are carefully tackled in all aspects of MATRYCS project. In the context of data sharing, privacy is strictly monitored to ensure compliance with the General Data Protection Regulation (GDPR) principles. More specifically, any personal data shared by the pilot partners throughout the MATRYCS project must be carefully inspected to ensure that any identifying attribute is anonymized. The entire process was thoroughly investigated through a security-specific questionnaire that was answered by all partners. Until now, personal data are being anonymized from the pilot partners side before shared to the consortium. However, the issue of sharing personal data will be carefully monitored throughout the project as more datasets are available.

Moreover, in the context of the end services designed in the context of the project, data privacy is also secured through the adoption of a security framework (WP3) that allows permission right only to users with the specified role, thus limiting any unauthorized users to access or modify data outside their access privileges. No personal data, as defined in D2.2 - "MATRYCS Technical & Security Specifications", (e.g., telephone numbers, building owners' identity) are included in the final services.

3.2.3 Toolbox

Generating business value from data is the main goal of every big data architecture, having a marketplace that facilitates the buying and selling of data and services is a major requirement to bear in mind. The marketplace can potentially span over multiple cloud-based deployment models:

- **Data as a Service (DaaS)**: it refers to the possibility for stakeholders to get access and visualize data on demand.
- **Software as a Service (SaaS)**: it refers to possibility for stakeholders to use specific software (business applications) on-demand to process data.
- **Platform as a Service (PaaS)**: it refers to the possibility for stakeholders to use a remote computing platform to provision, instantiate, run and manage services and applications.
- **Infrastructure as a Service (IaaS)**: it refers to the possibility for stakeholders to use a remote virtualized infrastructure (managed with cloud orchestration) to provision processing, storage and other computing resources.

The functional domain of the architecture must foresee the presence of such a marketplace and, accordingly, of all the software components, tools and services required for handling the marketplace and for logging transactions, offering assets, exploring available resources, providing clearing mechanisms and billing functions, etc.

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

4 MATRYCS Reference Architecture

The final version of the MATRYCS Reference Architecture is presented in this section by taking into consideration the requirements and the specifications of the previous sections. In the D2.3, the MATRYCS Reference Architecture is directly mapped to the Big Data Value Chain, which reflects the goal of providing a framework to support the creation of data value. The final version of the MATRYCS Reference Architecture following the same three layers: MATRYCS Governance, MATRYCS Processing and Analytics.

In particular, the final MATRYCS Reference Architecture mostly focuses on the functional/implementation viewpoint of the architectural model and it describes the set of basic components needed for the creation of the overall building big data ecosystem and of a building sector-related data economy. A functional view of the MATRYCS Reference Architecture is illustrated in Figure 4. The MATRYCS Reference Architecture in the view of technology and tool is shown in Figure 5. The rest of this Section provide more details regarding the software components included at each layer of the architecture.

4.1 General Technical Specifications

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 satisfied from the MATRYCS Reference Architecture. 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.

In this section, the final iteration for the finalisation of the Technical Specification is presented. In the below table, MATRYCS final general Technical Specification along with related User Requirements from which each specification has been elicited.

Table 6: MATRYCS General Technical Specifications

ID	Technical Specification 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
TR052	The MATRYCS platform must provide the possibility for future functionalities, ensuring system scalability and extensibility.	NFR029, NFR030
TR053	The MATRYCS platform will follow an incremental approach which facilitates the design and planning off the platform's functionalities through an iterative	NFR029, NFR030

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

	process	
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

Finally, a total of 56 Technical Specifications were elicited with 8 general ones and 48 specific ones per technical component. The Technical Specifications for MATRYCS GOVERNANCE, MATRYCS PROCESSING and MATRYCS ANALYTICS are 19, 14 and 15 respectively. The component related MATRYCS Technical Specifications are presented per component of the MATRYCS GOVERNANCE, MATRYCS PROCESSING and MATRYCS ANALYTICS layers, in the respective sections following the same format and referring the User requirements which are satisfying.

All the tables with the final Technical Specifications derive from 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 MATRYCS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.101000158

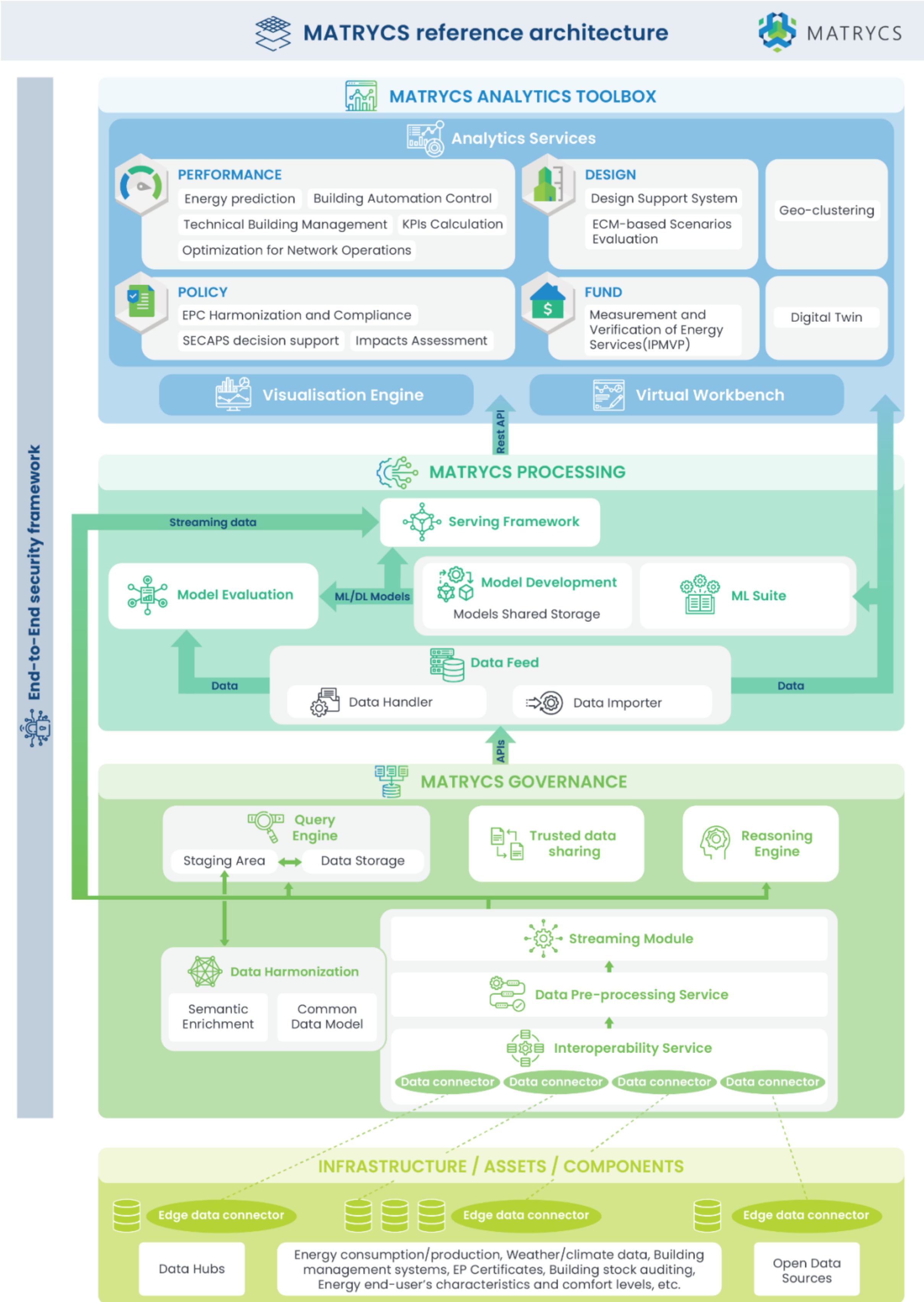


Figure 4 Functional View of the MATRYCS Reference Architecture

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

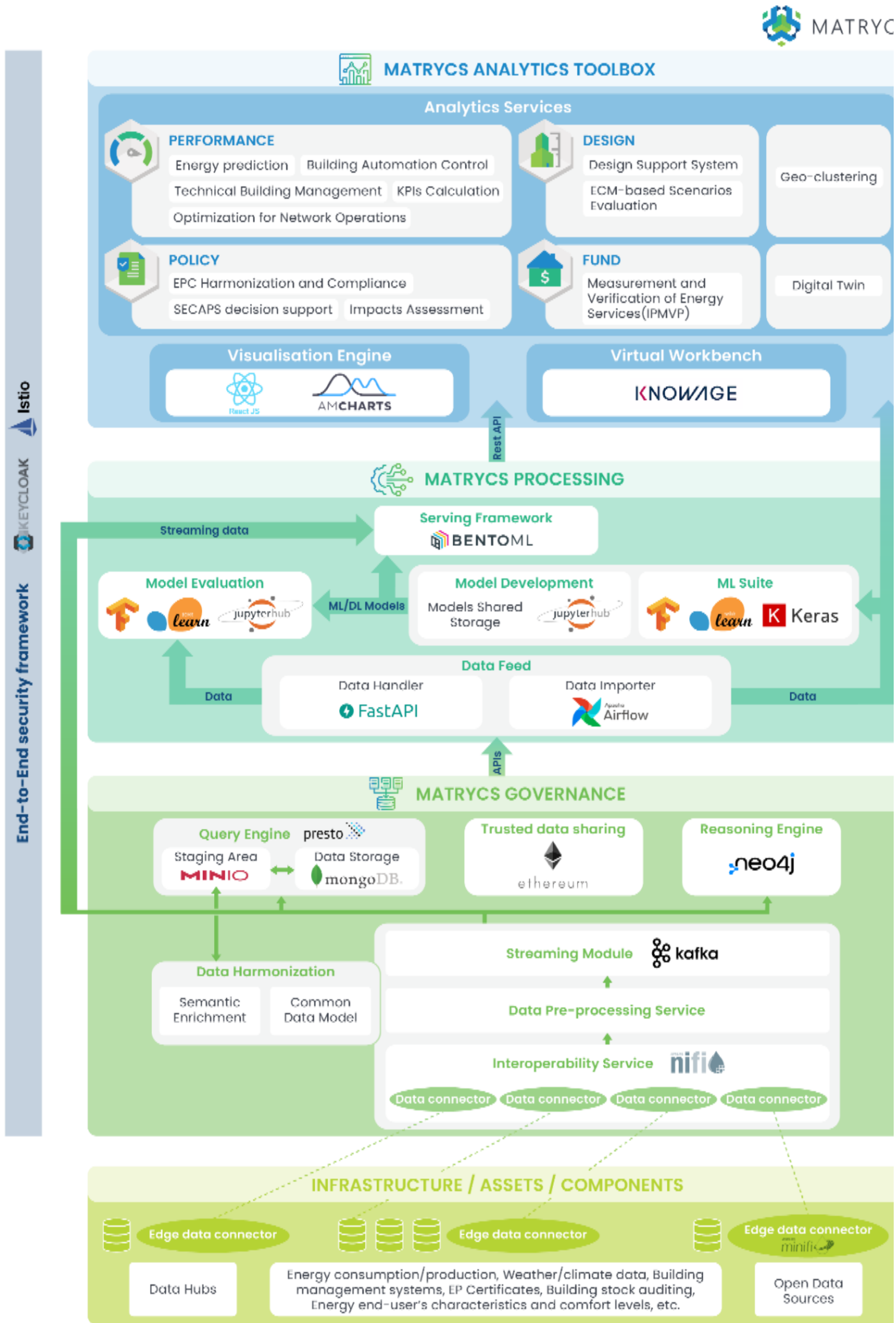


Figure 5: Technology View of the MATRYCS Reference Architecture

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

4.2 MATRYCS Governance Layer

The main objective of the MATRYCS Governance Layer is to provide the necessary middleware (see Figure 6) to act as a mediator between MATRYCS data providers and MATRYCS data users (analysis tools and services). Adopting a holistic approach to data management, the MATRYCS Governance layer takes into account different domain and non-domain data such as building data, energy data, sensors data, energy usage data (heating, lighting, cooling, air conditioning, ventilation), weather data, etc., and provide interoperability across this heterogeneous set of data via harmonization services and semantic enrichment.

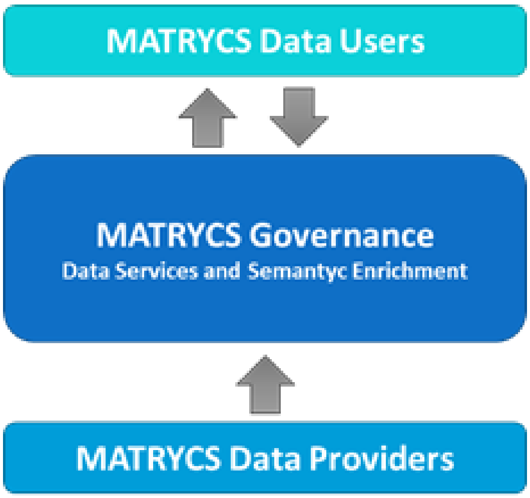


Figure 6: MATRYCS Governance overview

The MATRYCS Governance Layer provides data services that allow the integration, pre-processing, semantic annotation and querying of heterogeneous data, guarantying traceability, provenance tracking and accountability of the MATRYCS data. In this regard, the MATRYCS Governance architecture has been designed identifying the key high-level big data management phases needed to cover the whole MATRYCS Governance data pipeline processes (see Figure 7).

These processes include:

- **Data acquisition**: it is guaranteed via the *Interoperability Service Module*, which is designed to facilitate the data integration of heterogeneous data sources and/or platforms belonging to different MATRYCS actors. A *Streaming module* is in place to manage the data streaming to the MATRYCS Processing layer and for the in-memory processing of the low latency near real-time data.
- **Data Curation**: it is guaranteed via the *Data Pre-processing Module*, which is in charge of the data pre-processing activities for data cleansing, data curation, data anonymization and semantic annotation. This module also covers data modelling activities leveraging on pre-existing vocabularies and ontologies with the aim to define the MATRYCS Common Data Model.

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

- **Data storage**: it is guaranteed through two different modules: the *High Performance Query Engine* and the *Reasoning Engine*. The *High Performance Query Engine* builds on top of a NoSQL data storage and has the aim to perform complex queries in very efficient and highly scalable way. The *Reasoning Engine* is able to persist semantic datasets and any Resource Description Framework (RDF) information produced by the *Pre-processing & Semantic Enrichment Module*. Moreover, it can also consume JSON data and be used to construct the graph entities and relationships. Both modules expose intelligent querying systems and APIs to be used both in the MATRYCS Processing and in the MATRYCS Analytics layers.

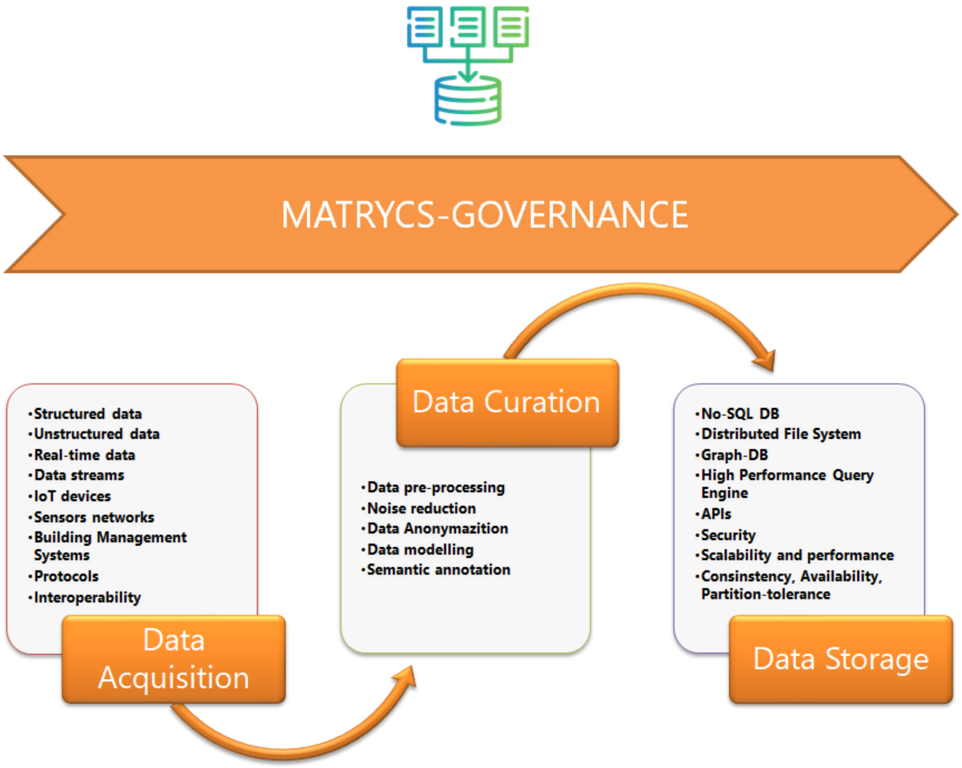


Figure 7: Big Data Management Phases in the MATRYCS Governance Layer

4.2.1 Interoperability Module

4.2.1.1 Design and functionality

The *Interoperability Service Module* is in charge of integrating and sharing of the heterogeneous data belonging to the different MATRYCS Data Providers, as well as external services, in an interoperable manner. This accounts for the fact that data ingested via the Interoperability Module can potentially come from a variety of data sources, such as Building Management Systems, Building Information Models, third-party platforms, sensors, IoT devices, open data repositories, etc.

The interoperability module will be designed to provide a consistent and reliable system, that can acquire a large amount of data and process different datasets in parallel. Specific goals for the module will include implementation of reusable components and configurations, automation of the data flow

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

and seamless integration of all the datasets. It will have to be powerful enough to support all the different data sources from all the LSPs and allow high data throughput. It will also need to send data to the Streaming module in a fault-tolerant way.

This module will include microservices (data connectors) to handle the different data sources provided with different data format and different communication protocols (e.g., Rest APIs, SFTP, IoT protocols, Sensor Network) providing also interfaces to other third-party systems and external datasets/platforms. This module will also be able to cover data exchange mechanisms at edge level, with the aim to offer both local computational power and storage capacity to services and functions that may need ultra-low latency, as well as local processing without leveraging on core cloud remote services.

4.2.1.2 Technologies and tools

After the detailed research and comparison of several platforms, Apache NiFi is recognised as the most suitable technology that can manage all the major aspects of the Interoperability module:

- NiFi is an open-source technology under the Apache License 2.0.
- It automates the data flow between systems.
- NiFi can acquire data from REST APIs, different databases, SFTP servers.
- Apache NiFi provides an easy-to-use graphical web interface to automate, run, and coordinate workflows.

The usage of MiNiFi is also envisioned, which can help to improve the efficiency of the whole environment. It has a lightweight processor and low usage of resources. Operations done on edge level can help to better the performance of the module. By gathering data at the source of creation and integrating with NiFi, MiNiFi will enable more effective dataflow and data management.

The SSH File Transfer Protocol (SFTP) servers are created as the endpoints to data upload by pilot sites and data retrieval by Apache NiFi. Each pilot site has a dedicated SFTP server for the upload of datasets with the agreed frequency, as NiFi has processors which can easily acquire and process data from SFTP folders.

State-of-art technologies are used for the implementation of the Interoperability module. The module is created by combining the right technologies and integrating them. Usage of the Apache NiFi so far leads to the conclusion that this tool enables high configurability, reliable data gathering, good parallelization of the workload, low overheads, data provenance, effective integration of data and scalability options.

4.2.1.3 Technical Specifications met

Table 7 MATRYCS Interoperability Module Technical Specifications

ID	Description of Technical Requirement satisfied	Related User Requirements
TR009	The MATRYCS platform must integrate cross-platform heterogenous data.	FR028, NFR009

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

TR010	The MATRYCS platform must collect real time streaming data.	NFR007
TR011	The MATRYCS platform must provide interoperability with other systems and platforms.	NFR008, NFR010
TR017	The MATRYCS platform must support importing datasets in different formats (csv,xlsx, etc.).	FR027, FR035, FR036, FR041
TR018	The MATRYCS platform must support importing large files.	FR027

4.2.2 Data Pre-processing Module

4.2.2.1 Design and functionality

The Data Pre-processing Module is composed of two main sub-components, namely the Data Cleansing & Anonymization Module and the Semantic Enrichment Module.

The goal of the Data Cleansing & Anonymization Module is to enhance the quality of the data and to guarantee privacy protection. It aims to improve the data quality, dedicated functions perform data cleansing, which include operations such as data restructuring, predefined value substitution, noise reduction, outlier detection, and data inconsistencies handling. All data are then transformed into the MATRYCS Data Model, which ensures all the data processed by the system adhere to the predefined standardized vocabularies and data models. In addition, to protect privacy, sensitive information is detected and anonymized in this module. The Data Cleansing & Anonymization Module is connected to the Data Steaming Module, which manages the data steaming towards different consumers.

The goal of the Semantic Enrichment Module is to overcome the heterogeneity of data and to guarantee semantic interoperability. The semantic enrichment process relies upon predefined ontologies, which cover endogenous buildings technical measurements as well as other context data, such as weather, geographical information, or energy networks data. Ontologies contain the vocabulary and the set of possible relationships, which follow well-established standards. Based on those vocabulary and relationships, it is possible to guarantee the semantic interoperability needed at application level. The Semantic Enrichment Module publishes the semantically annotated data through the Data Steaming Module.

4.2.2.2 Technologies and tools

In the Data Cleansing part, different outlier detection technologies are used. K-means clustering is a vector quantization approach that seeks to split n observations into k clusters, with each observation belonging to the cluster with the nearest mean, which serves as the cluster's prototype. Based on the clustering results, the smallest cluster will have the highest possibility to be an outlier.

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

For Anonymization part, two step are required. First, personal data detection method should be developed. Afterwards, the sensitive information should be removed or masked. The different approach is shown as following:

- k-anonymity
- l-diversity
- neighbourhood attacks
- k-degree anonymity
- k-isomorphic neighbour anonymity
- k-automorphism

The last part is Semantic Enrichment Module. The python library RDFlib is used. RDFlib contains the parsers and serialization of different ontology data format. It provides the graph interface to a single graph, multiple named graphs, and a dataset of graphs.

4.2.2.3 Technical Specifications met

Table 8 Data Pre-processing Module Technical Specification

ID	Description of Technical Requirement satisfied	Related User Requirements
TR001	The MATRYCS platform must provide data cleansing functionalities.	FR031
TR002	The MATRYCS platform must provide a data anonymisation tool / service.	FR030 deleted, no mapping UR
TR003	The MATRYCS platform must comply with a common data model.	FR045
TR005	The MATRYCS platform must ensure that external data being imported in MATRYCS are mapped to the MATRYCS data model.	NFR009, NFR021
TR012	The MATRYCS platform must provide semantic annotation of data.	NFR009

4.2.3 Data Streaming Module

4.2.3.1 Design and functionality

The Data Streaming Module is a key component within the MATRYCS Governance Layer, as it allows re-routing the data to the different software components. It manages the stream of MATRYCS data, handling dynamically the frequency rate of the data streaming for the subsequent in-memory processing of the low latency near real time data.

The Streaming module is designed to act as a unified framework to manage real-time data from all MATRYCS data providers. It needs to have a high message throughput to handle large volumes of data and low-latency delivery to handle several consumers. In this way, the Streaming module can operate as a traditional messaging platform and as an event streams system, ready for distributed real-time data processing. The module also ensures fault tolerance and efficiency by employing clustered

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

solutions.

4.2.3.2 Technologies and tools

There are many types of data streaming platforms, but based on the research carried out, Apache Kafka is selected to be used as the core of the Streaming module in the MATRYCS project. Apache Kafka and all its sub-components are available as open source under the Apache License v2.0 license. Kafka works as a data streaming pipeline. It is highly configurable, parameters such as offsets and partitions are set to create an appropriate environment for the purposes of the project. The effective methods for the partition and the multi-thread usage are studied to improve efficiency in the module.

Furthermore, a set of additional open-source tools from the Confluent Platform is deployed. These tools give means to better manage and organise the data operations in the Streaming module. Confluent open-source tools include several different connectors, Schema registry for metadata management, and REST Proxy with several endpoints.

Kafka is a message queue tool and can provide real-time data transfer. It uses publish/subscribe model and a concept of topics. Each dataset has its own topic, so consumers can choose which dataset they want to be subscribed to. Moreover, Apache Kafka is an event streaming platform, enabling continuous publish/consumption of the data. On the other hand, it supports the heavy production of messages when historical data are sent to the Streaming module.

Kafka enables in-memory and disk-based fast storage, and a configurable maximum size of messages. Data can be preserved in the Kafka Broker for a customizable amount of time, seven days by default. As Kafka persists messages, storage volumes are attached to the Kafka broker instances. Low-latency and high-throughput of messages are also implemented in Kafka. Apache Kafka provides real-time data analysis and operational monitoring, which the connected MATRYCS modules can use to have up-to-date information.

Usage of Kafka cluster configured and deployed in the described manner leads to better performances and efficient data streaming in a secure environment. Kafka's functionalities set a distributed, fault-tolerant, scalable, and secure system. The specified deployment gives an automated and maintainable environment for all tools involved.

4.2.3.3 Technical Specifications met

Table 9 Data Streaming Module Technical Specification

ID	Description of Technical Requirement satisfied	Related User Requirements
TR010	The MATRYCS platform must collect real time streaming data.	NFR007

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

4.2.4 Data Storage & Query Engine

4.2.4.1 Design and functionality

The MATRYCS Data Storage is consisted of 2 sub-components, the MATRYCS Storage consumer and the MATRYCS Staging Area. The MATRYCS Storage Consumer receives JSON events from Streaming module and persists them in Staging Area. It is a Python 3.8 Confluent Kafka Consumer that manages the data and stores them to MinIO LSP data bucket. JSON events from different topics are persisted to different sub-directories of the bucket. The following figure demonstrates the MATRYCS Storage Architecture.

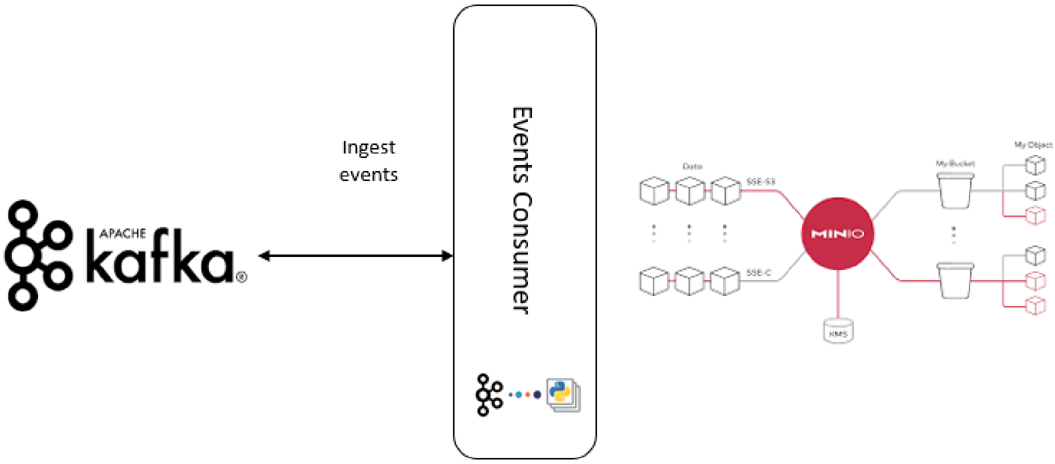


Figure 8: Data Storage Consumer

Part of MATRYCS Storage is also a MongoDB instance which is installed for being used from Data Feed Module and from MATRYCS PROCESSING, which will leverage data feed module to receive batch of data from MinIO LSP bucket for further pre-processing in order to enhance models input in training phase of MATRYCS.

The Query Engine is a Presto Instance installed on top of Staging Area and on MongoDB in order to have a database agnostic engine. MATRYCS upper layers, ANALYTICS & PROCESSING, will leverage SQL for interacting with processed and unprocessed data.

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

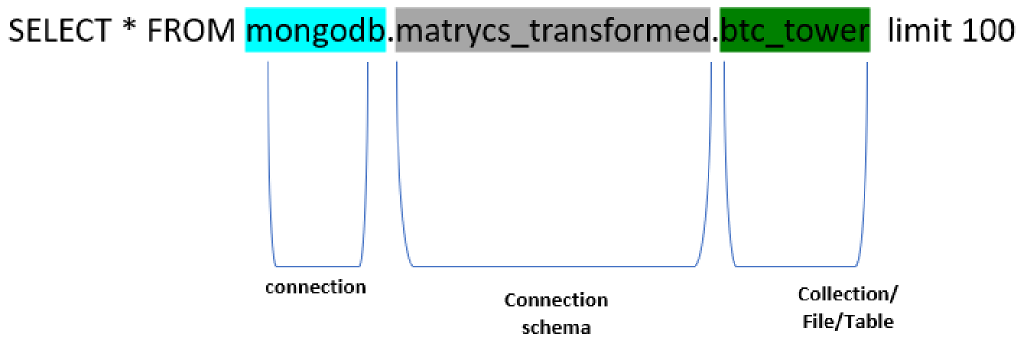


Figure 9: Query in MATRYCS Query Engine

4.2.4.2 Technologies and tools

Table 10 Data Storage & Query Engine Technologies and Tools used

Technology	Description
Python 3.8	Programming Language used
MinIO	Staging Area used for MATRYCS GOVERNANCE
MongoDB	NoSQL Data Storage
Confluent Kafka Consumer	Consumer of streaming events
Docker	Deployment approach

4.2.4.3 Technical Specifications met

Table 11 Data Storage & Query Engine Technical Specifications

ID	Description of Technical Requirement satisfied	Related User Requirements
TR004	The MATRYCS platform must provide secure data storage functionality.	NFR004, NFR013
TR006	The MATRYCS platform must provide a high-performance distributed query engine.	FR 037, NFR003, NFR014
TR007	The MATRYCS platform storage system must provide optimal space allocation.	NFR004

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

TR016	The MATRYCS platform must offer a well-defined API for data export.	FR029

4.2.5 Reasoning Engine

Reasoning Engine is a component that provides intelligent querying, insights and search capabilities by leveraging the available knowledge for digital twins and other building analytics services. Furthermore, it provides functionalities for adding more information to the existing knowledge base exploiting metadata and relationships among data. The Reasoning Engine must be capable of consuming data that received from the **Semantic Enrichment Module**, via the **Streaming Module**. These data could be expressed with different data models, including metadata data models (such as RDF) and when received they are persisted to the **Reasoning Engine's** graph database. The graph database is a powerful inference engine that enables graph functionalities over entities and allows connections for extracting new insights and patterns from available datasets.

4.2.5.1 Design and functionality

The Reasoning Engine Architecture is presented in Figure 10.

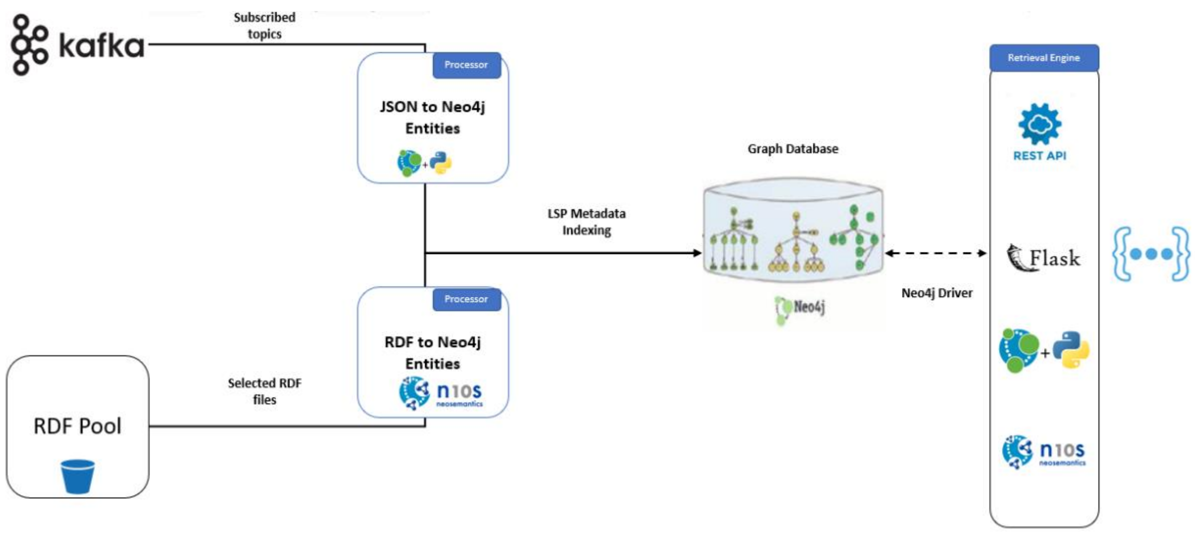


Figure 10: Reasoning Engine Architecture

Reasoning Engine receives metadata in order to create logical entities. These metadata are pipelined either from Kafka as JSON objects, or as RDF files stored in MATRYCS Staging Area. While JSON objects can be directly be imported in Neo4j graph database as a built-in feature provided, RDF files cannot be imported directly, Thus, the RDF files are transformed into Neo4j Entities via Python 3 and the Neosemantics plugin, before importing the metadata to Neo4j. Then, Rest APIs expose the stored graph information to MATRYCS upper layers.

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

4.2.5.2 Technologies and tools

Table 12 contains the technologies and tools used in Reasoning Engine's implementation.

Table 12: Reasoning Engine Technologies and Tools used

Technologies / Tools (version)	Description
Python 3.8	Programming Language
Neo4j: 4.0	Graph Database
Confluent Kafka: 2.3.1	Kafka library used for consuming JSON events from Streaming Module
Neosemantics: 4.0	Neo4j plugin which transforms the imported RDF files into Neo4j entities
Flask: 1.1.12	Python library used for creating REST APIs

4.2.5.3 Technical Specifications met

Table 13 Reasoning Engine Technical Specifications

ID	Description of Technical Requirement satisfied	Related User Requirements
TR013	The MATRYCS platform must offer a reasoning engine.	FR045

4.2.6 Blockchain/Distributed Ledger system

The goal of this module is the securitization of data storage, specifically the primary purpose of Blockchain technologies is to remove the need for intermediaries and replace them with a distributed network of digital users who work in partnership to verify transactions and safeguard the integrity of the ledger.

4.2.6.1 Design and functionality

Use of Blockchain in the data-sharing model has the following main key features: it ensures the traceability and data storage throughout a decentralised and distributed system, with which it will be possible to implement a secure way to track changes in information over time. In this way, the module can warrant the creation of trust among untrusted participants; in addition, the absence of intermediaries can promote a more transparent data sharing. In order to implement these functionalities this module will utilise Blockchain technology. The Blockchain is a distributed ledger, based on a shared and distributed database, containing a log of transactions in chronological order. Transactions are grouped into blocks and chained through cryptographic hashes into an ongoing chain of hash-based proof-of-work, forming a record that cannot be changed without redoing the proof-of-work.

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

Within the MATRYCS Governance Layer, this module interacts with the Data Storage for the extraction of harmonised data and with the Blockchain network for the physical fingerprints (hash codes) storage in the chain blocks to assure data immutability. This will be the main mechanism to assure immutability of data stored into the Governance Data Layer; optionally, the module could also interact directly with the MATRYCS assets, and external data providers; in this case, the interaction will require specific data connectors to integrate the datasets with different standard and access interfaces.

4.2.6.2 Technologies and tools

Blockchains can be distinguished according to the consensus mechanism and programming capabilities. Considering the consensus mechanism, Blockchains differ in the definition of the participation of nodes in the distributed network and the roles that they can perform. In particular, open Blockchains and permissioned (or private) Blockchains can be distinguished. Considering the programming capabilities, it is possible to differentiate between Blockchains programmable via simple scripting and Blockchains providing Turing-complete computational capabilities, enabling the creation of "smart contracts". Ethereum was the first Blockchain supporting smart contracts and it is still the most notable example of Turing-complete programmable Blockchain.

Ethereum was proposed by Vitalik Buterin in 2013 (Ethereum Whitepaper) and further detailed by Gavin Wood in the 'yellow paper' (Ethereum: A Secure Decentralised Generalised Transaction Ledger (Wood)) . As the Ethereum website (Ethereum) reports, "Ethereum is a decentralized platform that runs smart contracts." These contracts run on the "Ethereum Virtual Machine" (EVM), a distributed computing network made up of all the devices running Ethereum nodes. Like other blockchains, Ethereum has a native cryptocurrency called Ether (ETH) and it has a double use: it is used as an incentive for the network "validators", but also to regulate the use of the blockchain computational resources.

4.2.6.3 Technical Specifications met

Table 14 Blockchain/ Distributed Ledger System Technical Specifications

ID	Description of Technical Requirement satisfied	Related User Requirements
TR008	The MATRYCS platform must provide a blockchain platform to enable secure data and models sharing for cross-stakeholders.	NFR020

4.3 MATRYCS Processing Layer

MATRYCS Processing Layer is responsible for advanced data processing. It aims at encapsulating the intelligent building blocks of MATRYCS solution, by providing a library of reusable ML/DL models. The ML/DL models are made available with the view to promote their quick adaptation and reuse along different contexts. Figure 11 depicts the MATRYCS Processing Architecture.

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

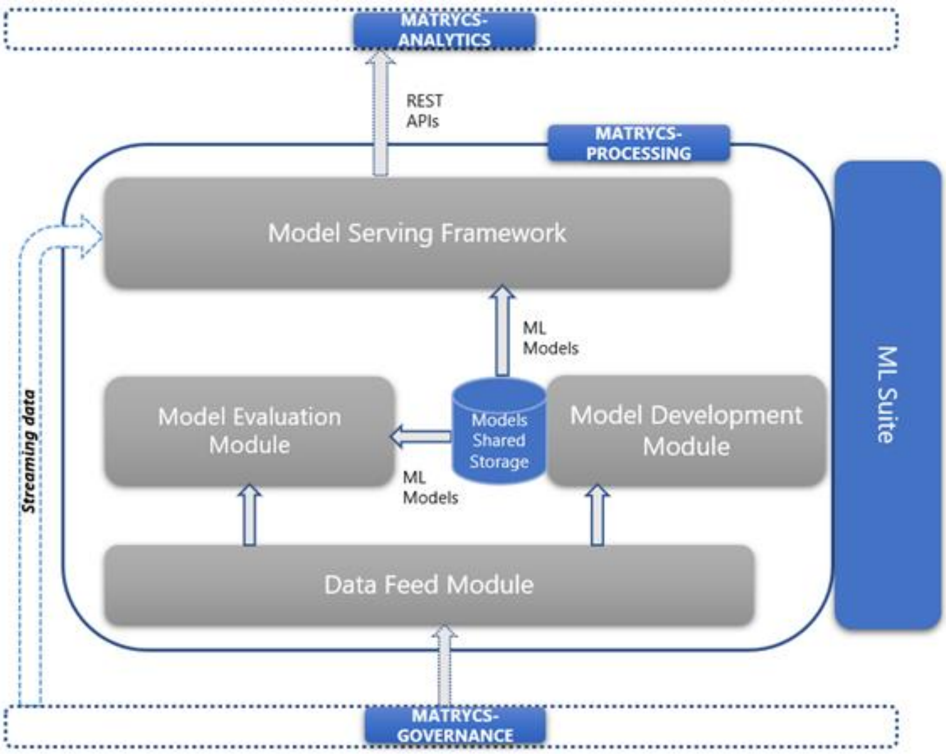


Figure 11: MATRYCS PROCESSING Architecture

The MATRYCS Processing Layer is connected to the MATRYCS Governance Layer through the **Data Feed Module** and the **Model Serving Framework**. The Data Feed Module is the mediator taking data from MATRYCS Governance Layer and makes it available into MATRYCS Processing Layer. Then data preparation, model training and the final evaluation of the AI-based services can be performed. ML/DL models successfully trained and validated are finally made available through the **Model Serving Framework**. The Model Serving Framework is the second connection point to MATRYCS Governance Layer, receiving streaming data from the **Streaming Module** and provides advanced analytics information, based on the results of the AI-based processing (such as forecasts, regressions, feature extraction, etc.). A short description of MATRYCS PROCESSING components and their connections details with components from MATRYCS GOVERNANCE are presented bellow (Table 15).

Table 15: MATRYCS PROCESSING COMPONENTS

Component	Connection with MATRYCS GOVERNANCE	Short Description
Data Feed Module	Connection with MATRYCS Storage & MATRYCS Query Engine	Component responsible for inserting harmonised data to MATRYCS PROCESSING. Acts as a sharing component and also prepares the data for Training
ML Suite	No Connection	Library of ML / DL Frameworks

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

Model Development Module	No Connection	The component which hosts the ML / DL Training of MATRYCS LSP datasets.
Model Evaluation	No Connection	Anaconda Environment used for trained models' evaluation.
Serving Framework	Connection with MATRYCS Streaming Module	Component that serves trained models as docker containers and expose their prediction functionality to MATRYCS ANALYTICS

Table 16 presents the core frameworks used in WP4 and MATRYCS PROCESSING Layer and in which components their functionalities are enabled.

Table 16: MATRYCS PROCESSING Core Frameworks

Framework	Component used	Description
Apache Airflow	Data Feed Module	Orchestrator used for inserting data to MATRYCS PROCESSING
PRESTO Client	Data Feed Module	Client Used for interacting with Presto Query Engine
MongoDB	Data Feed Module	MongoDB NoSQL database used as Data Feed Module storage
Keras / Tensorflow	Machine Learning Suite / Models Development	Python libraries used for Deep Learning models development
Scikit-Learn	Machine Learning Suite / Models Development	Python library used for Machine Learning models development.
MLflow	Machine Learning Suite / Models Development	Component Used for MATRYCS PROCESSING Models Development
Anaconda Environment	Model Evaluation	Environment used for Model Evaluation
BentoML Serving Framework	Serving Framework	BentoML Python Framework for serving trained models

4.3.1 Data Feed Module

Data Feed Module is consisted of 2 sub-components, the Data Importer and the Data Handler.

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

Data Importer is responsible for receiving the data from MATRYCS STAGING area, then conducts extra pre-processing steps (i.e., min-max scaling, one hot encoding, drop duplicates and null variables) and stores the prepared data into MATRYCS MongoDB. Apache Airflow and its schedulers orchestrate periodical tasks to pick up stored data from Staging area.

Data Handler is the collection of REST APIs that receive the prepared LSP data from MATRYCS MongoDB and distributes them towards other MATRYCS PROCESSING components and MATRYCS ANALYTICS layer.

4.3.1.1 Design and functionality (WP4)

Data Feed Module receives data from MATRYCS GOVERNANCE and to do so, the Staging Area is being leveraged for picking harmonised LSP data. Then basic procedures are conducting from Data Importer and finally the data are stored in MongoDB collections. Data Handler is the collection of REST APIs that are used for aggregations and query pre-processing, by leveraging PRESTOSQL and Pandas DataFrames. Figure 12 depicts the module's architecture and the interactions with MATRYCS-GOVERNANCE layer are highlighted.

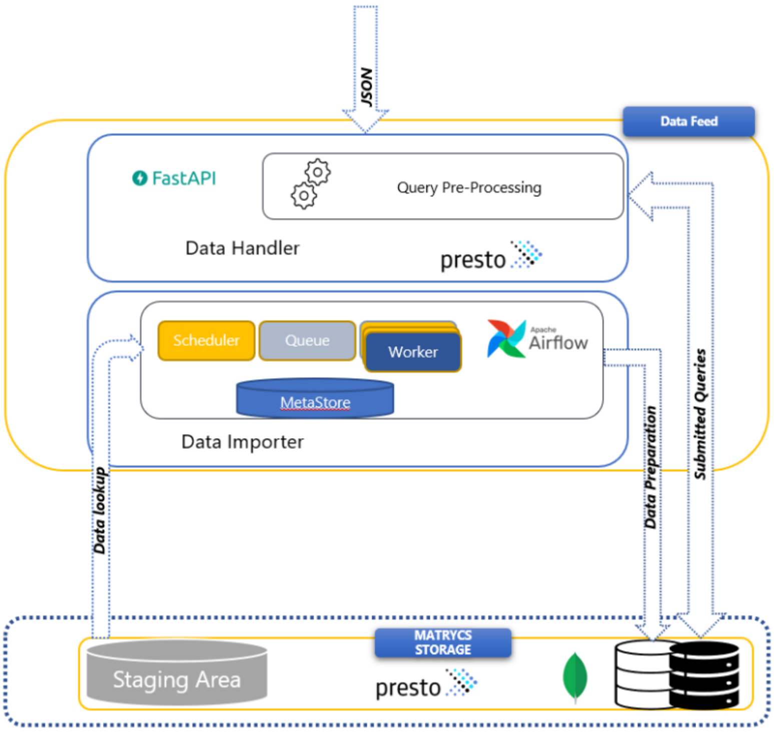


Figure 12: Data Feed Module Architecture

4.3.1.2 Technologies and tools

Table 17 presents the technologies and tools used in Data Feed Module implementation.

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

Table 17: Tools & Technologies used in Data Feed Module implementation

Technologies / Tools	Description
Apache Airflow	Orchestrator mechanism used for periodical inserts in MATRYCS PROCESSING
MongoDB	Storage for inserted / aggregated data in MATRYCS PROCESSING
Presto Client	Client for interaction with Presto
Pandas DataFrames	Library used for DataFrames management
PyMongo	Python library used for interaction with MongoDB database
FastAPI	Python library used for REST API development

4.3.1.3 Technical Specifications met

Table 18 Data Feed Module Technical Specifications

ID	Description of Technical Requirement satisfied	Related User Requirements
TR016	The MATRYCS platform must offer a well-defined API for data export.	FR029
TR043	The MATRYCS platform must allow a user to easily perform aggregations on a dataset.	FR012, FR032

4.3.2 Machine Learning Suite

For the *Machine Learning Suite*, we designed and developed a configurable general-purpose ML pipeline to train data-driven ML models. Considering data provided by *Data Feed Module* (Section 4.2.1) from different project partners, the suite focuses on tabular data type. The suite supports three major types of tasks deferred to ML by design. These are classification, regression, and clustering tasks. In addition, we support automated parameter tuning of the selected ML algorithms.

The Machine Learning Suite of MATRYCS PROCESSING is being used from Model Development module, that hosts the training phase of MATRYCS LSP data, for training the algorithms using different technologies, frameworks and programming languages.

4.3.2.1 Design and functionality

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

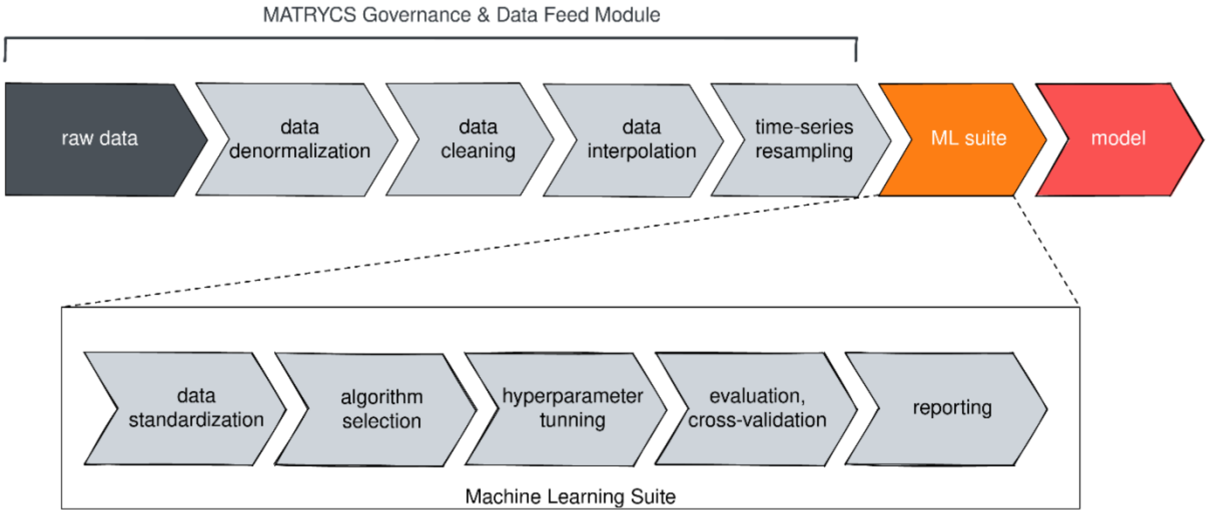


Figure 13 Machine Learning Suite.

The Machine learning Suite is designed as a parametrized pipeline. Its pipeline design allows being easily extended or improved based on the project requirements. At the time of reporting, the *Machine Learning Suite* supports the following functionalities:

- Can request and ingest data from *Data Feed Module* (Section 4.2.1)
- Supports three major types tasks deferred to ML: **classification**, **regression**, and unsupervised **clustering**.
- Supports a variety of algorithms per type of task: **Classification:** Logistic Regression, Decision Tree, SVM classifier, XGBoost Classifier **Regression:** Linear regression, Decision trees, SVM regressor, XGBoost Regressor **Clustering:** K-Means, DBSCAN
- Supported algorithms can use either default parameters or automated parameter tunning using grid search or AutoML.
- Extensive 10-Fold cross-validation to ensure fairness and consistent comparison
- The evaluation step considers multiple performance metrics
- The progress, final model, and evaluation report are automatically submitted to the *Model Development Module* (Section 4.2.3)

4.3.2.2 Technologies and tools

For the *Machine Learning Suite*, we used exclusively open-source tools and technologies. We used the following:

- **Python** programming language because of its large community and its extensive repository of production-ready libraries and frameworks for development.
- **Scikit-learn** framework with one of the richest and well-documented portfolios of data transformation functions, validation techniques, metrics, pre-processing, and machine learning algorithms. Also, it is backed by one of the largest ML communities.
- **Matplotlib** is one of the well-known data visualization libraries. Here, we use it to visualize metrics, losses, and statistical error distribution at the reporting phase.

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

- **MLFlow client** is used for submitting model(s), experiment(s), and its report(s) to centralised logging service presented in *Model Development Module* described (Section 4.2.3)

```python
    with mlflow.start_run(run_name='linear-regression') as run:
        model = linear_model.LinearRegression()
        model.fit(X_train, y_train)

        y_pred = model.predict(X_test)

        ## Metrics
        rmse = metrics.mean_squared_error(y_true, y_pred, squared=False)
        mlflow.log_metric('RMSE', rmse)

        mae = metrics.mean_absolute_error(y_true, y_pred)
        mlflow.log_metric('MAE', mae)

        r2 = metrics.r2_score(y_true, y_pred)
        mlflow.log_metric('R2', r2)


        ## Figures / Plots
        if hasattr(model, 'feature_importances_'):
            # get importance
            importance = model.feature_importances_
            fig, ax = plt.subplots()
            # plot feature importance
            ax.bar([x for x in range(len(importance))], importance)
            mlflow.log_figure(fig, artifact_file='feature_importance.png')

        # CDF
        n_bins = 50
        fig, ax = plt.subplots()
        x = np.abs(y_true - y_pred)
        n, bins, patches = ax.hist(x, n_bins, density=True, histtype='step', cumulative=True)
        ax.grid(True)
        #ax.legend(loc='right')
        ax.set_title('CDF')
        ax.set_xlabel('Estimation error')
        ax.set_ylabel('Likelihood')
        mlflow.log_figure(fig, artifact_file='cdf.png')
```

Figure 14 Machine Learning Suite Code Example

4.3.2.3 Technical Specifications met

Table 19 Machine Learning Suite Technical Specifications

ID	Description of Technical Requirement satisfied	Related User Requirements
TR032	The MATRYCS platform must provide tools and services to apply and evaluate machine learning algorithms	NFR006
TR033	The MATRYCS platform must provide tools and services to apply and evaluate deep learning algorithms	NFR006

4.3.3 Model Development Module

The purpose of the Model Development Module is to complement and interconnect the ML Suite (section 4.3.2), which is responsible for training data-driven models on top of data provided by

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

the Data Feed Module (Section 4.3.1), and Model Evaluation Framework (Section 4.3.4). In addition, the Model Development Module helps keep track record of data-driven model development progression, helps select the best candidate models, and promotes specific models to stagging and later to the production phase.

4.3.3.1 Design and functionality

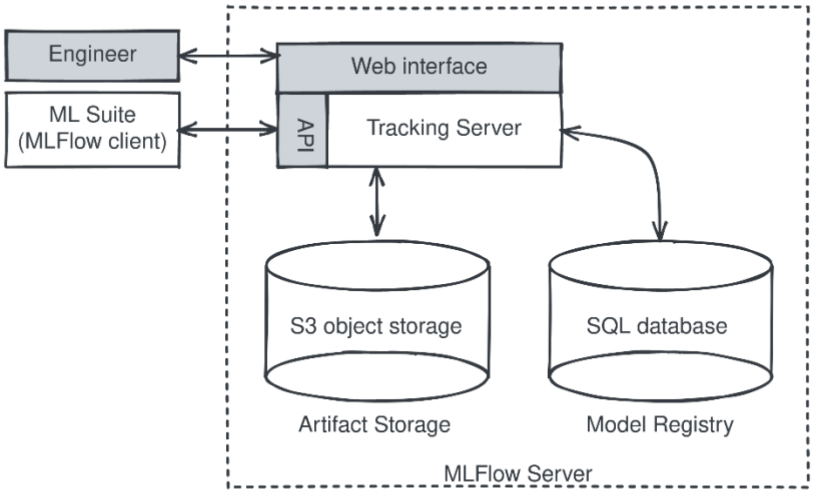


Figure 15: MLFlow server workflow

The design of the *Model Development Module* consists of the following services: **tracking server**, **model registry**, and **artifact storage** service. The ML Suite builds a data-driven ML model and submits it to the Model Development Module. The tracking server ensures that a dedicated "bucket" for storing binary data of the model is created on artifact storage. Models' metadata, bucket location, and evaluation metrics are recorded in the model registry. Since the MLFlow logs all the experiments, clients can recall/retrieve any registered model through the same API.

While the MLFlow server can be configured programmatically through API, for ease of use it offers an intuitive web frontend (see Figure 16) to track and visualize recorded experiments. Within few clicks, engineers can promote certain model to *stagging*, and later to *production* grade, if the model(s) pass the evaluation in the Model Evaluation Framework (Section 4.3.4).

For the Artifact Storage we used MinIO, which is an S3 API compliant object storage service. It stores all binary data related to the particular ML model, such as plots, model's internal state, and experiment requirements. MinIO also offers web frontend (see Figure 17) to inspect bucket content and simplify permission configuration (e.g., user/group creation, bucket permissions, bucket encryption).

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

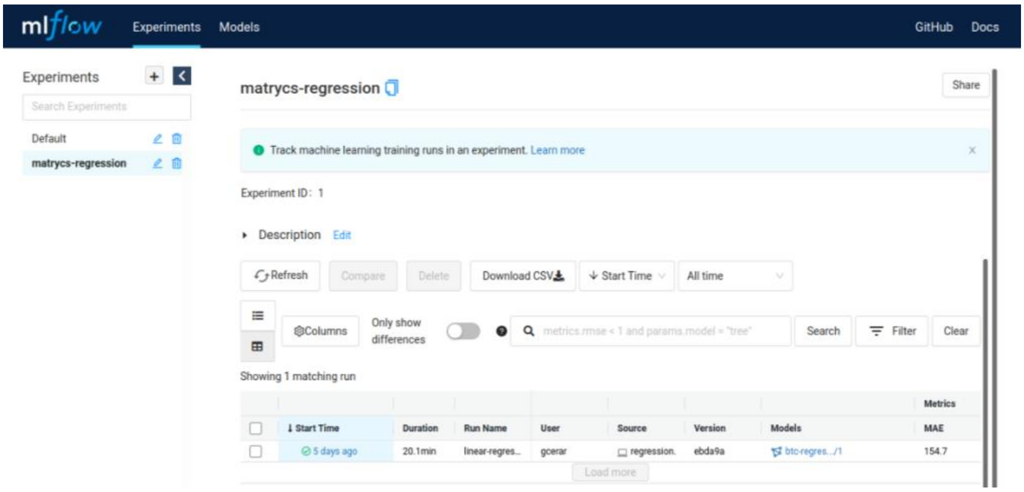


Figure 16: MLFlow's MLOps web frontend showing projects on left and list of experiments on the bottom.

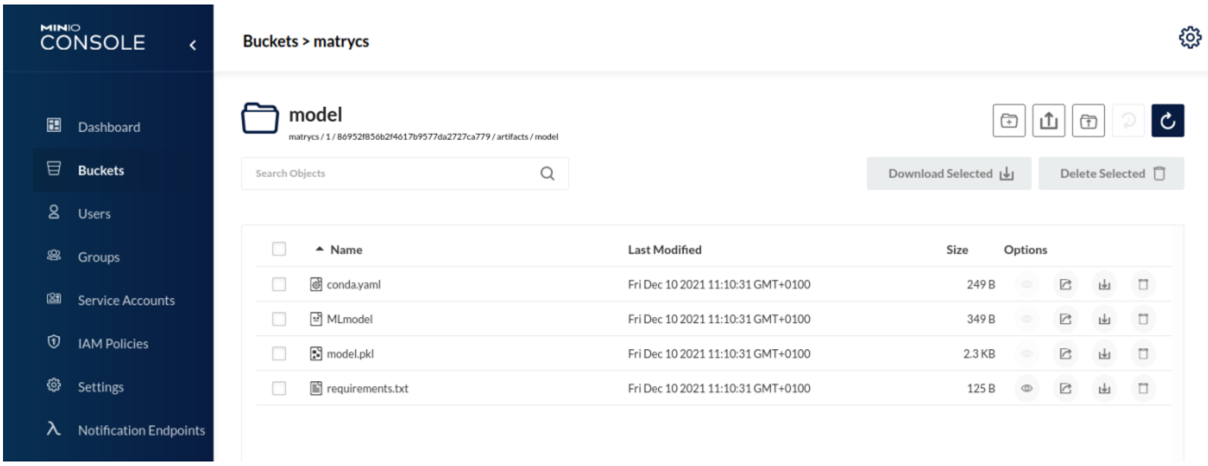


Figure 17: MinIO's web frontend. Shows content of one of the bucket containing ML model's artifact files such as requirements and binary ML algorithms' internal state.

4.3.3.2 Technologies and tools

The *Model Development Module* uses open-source solutions to build services and serve the task. We used the following tools:

- **MinIO** as an on-premise (self-hosted) service for object storage service (S3 compatible API) used for long-term storage of binary blobs, such as plots, (binary state of) ML models, and text description of software dependencies. Each model resides in a dedicated place called the "bucket".
- **Relational database** (i.e., PostgreSQL) is used for ML model registry. Each submitted model or experiment is registered into a relational database with model evaluation metrics for later comparison and analysis.

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

- **MLFlow server** is responsible for providing API for MLFlow clients to access. It provides an abstraction storing ML models, evaluation metrics, figures and other objects related to a particular model. Also, it provides an intuitive web frontend for easier management. For each submitted model to MLFlow, the server created a dedicated bucket (on MinIO) for binary blob and entry into ML model registry (i.e., relational database) along with a location to the newly created bucket.

4.3.3.3 Technical Specifications met

Table 20 Model Development Module Technical Specifications

ID	Description of Technical Requirement satisfied	Related User Requirements
TR032	The MATRYCS platform must provide tools and services to apply and evaluate machine learning algorithms	NFR006
TR033	The MATRYCS platform must provide tools and services to apply and evaluate deep learning algorithms	NFR006

4.3.4 Model Evaluation Framework

The *Model Evaluation Framework* is used for evaluating the trained models created within the *Model Development Module*. Therefore, it is interconnected with the *Models' Shared Storage* for getting the trained models and with the *Data Feed Module* for getting the dataset to be used for the evaluation purposes. These two inputs are combined for testing and evaluating specific metrics representative of the quality of the trained models. The assessment (and possible validation) of the trained models can be done exploiting metrics commonly adopted in the AI and ML/DL domain, such as the R^2 score, Mean Squared Error and Mean Absolute Error for forecasting and regression analysis, or the F1 Score, log loss/binary cross-entropy and categorical cross-entropy for classification problems. Using these metrics, the quality of the produced AI models can be ensured, therefore increasing the trustworthiness about the information created via such models.

4.3.4.1 Design and functionality

To design the Model Evaluation Framework the MLOps approach has been taken in consideration. It comes into play to unify the two processes of ML model development (Dev) which is the objective of this service module and commissioning (Ops) which is better specified in the section dedicated to the Model Serving Module reported in section 4.3.5. In particular, the model testing phase is the process that lies halfway between model development and training and the subsequent serving of the models. Figure Figure 18 shows the envisaged process starting from the data management, moving to the model developers and their evaluation, and concluding with the model serving.

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

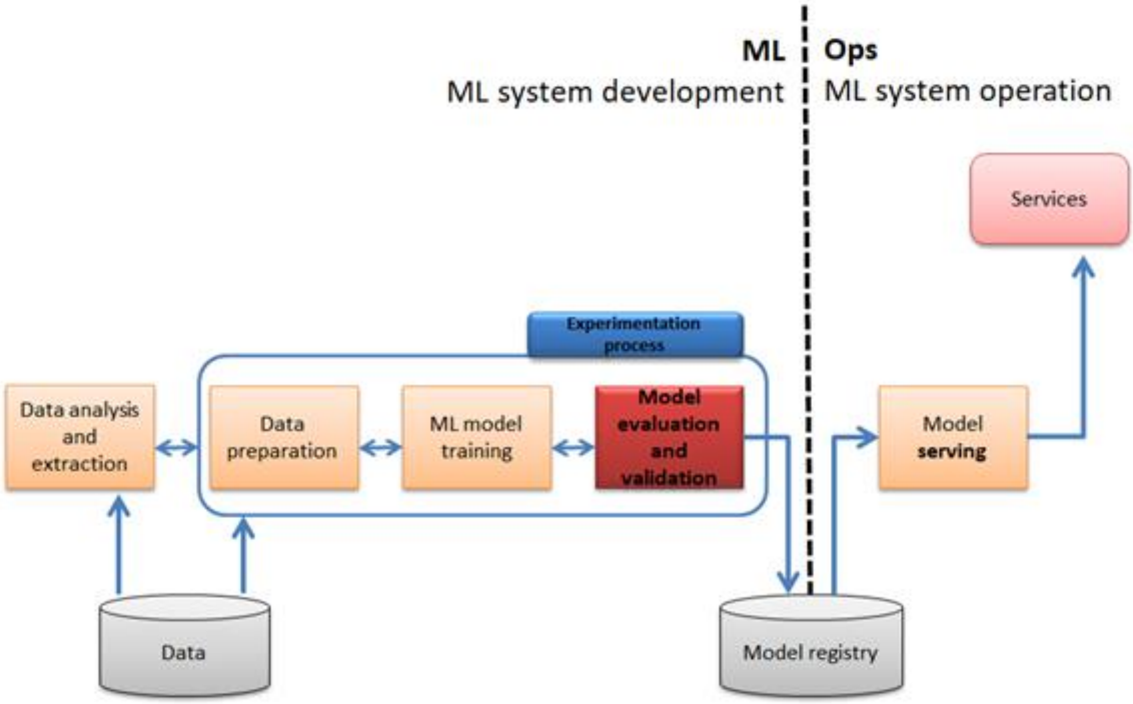


Figure 18 MLOps process – Model evaluation

The tools that compose the Model evaluation framework have been adopted to facilitate the model developers in the evaluation process of the developed models after their training phase. In particular, the main functionalities and features identified are the following:

- A common repository where to save the developed models and keep track of the different versions created during the finishing phase.
- A common access point to the different data sets of the project whether they come from assets of the different pilots or from sources outside the project.
- A common set of tools for testing and evaluating models and refining them if necessary.
- A common set of tools to keep track of the logs related to the processing results and related metrics that can also be easily analysed through a graphical interface.
- An environment capable of providing developers with the most common ML libraries and that can be easily updated with new ones if needed.

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

4.3.4.2 Technologies and tools

During the first phase of the project different technologies have been tested and adopted with the aim to provide a ready-to-use environment for evaluate the already trained ML models built within the context of the MATRYCS project. Special attention was given to have an environment and a series of tool potentially extendible and interoperable. The adopted solution is based on the Anaconda Data Science Platform[1]. This platform provides a set of tools to facilitate the data engineers and scientist in testing/update/evaluate/analyse their ML models.

More precisely, the following tools and technologies have been adopted, installed, and configured for the user needs:

- Anaconda[2] environment with Jupyter Notebook[3]
- Specific libraries that enable straightforward usage of trained ML models and related evaluation, such as Numpy[4], Pandas[5], TensorFlow[6], Keras[7], TensorBoard[8], Scikit-learn[9], Pytorch[10], Matplotlib[11], Seaborn[12], etc.
- Different Python libraries that enable connection to other modules and APIs such as Kafka Stream[13], Databases[14], Cloogy[15] API, etc.

4.3.4.3 Technical Specifications met

Table 21 Model Evaluation Framework Technical Specifications

ID	Description of Technical Requirement satisfied	Related User Requirements
TR032	The MATRYCS platform must provide tools and services to apply and evaluate machine learning algorithms	NFR006

[1] https://www.anaconda.com/
[2] Anaconda, https://www.anaconda.com/
[3] Jupyter, https://jupyter.org/
[4] Numpy, https://numpy.org/
[5] Pandas, https://pandas.pydata.org/
[6] Tensorflow, https://www.tensorflow.org/
[7] Keras, https://keras.io/
[8] TensorBoard, https://www.tensorflow.org/tensorboard
[9] Scikit-Learn, https://scikit-learn.org/stable/
[10] Pytorch, https://pytorch.org/
[11] MatplotLib, https://matplotlib.org/
[12] Seaborn, https://seaborn.pydata.org/
[13] Kafka Stream, https://docs.confluent.io/platform/current/streams/index.html
[14] Databases, https://pypi.org/project/databases/
[15] Cloogy, https://pypi.org/project/cloogy/

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

TR033	The MATRYCS platform must provide tools and services to apply and evaluate deep learning algorithms	NFR006

4.3.5 Model Serving Framework

The *Model Serving Framework* is the bridge between the MATRYCS Processing and the MATRYCS Analytics layers. It serves the ML/DL models saved under the trained models library (hence those models already trained by the ML developers and subsequently evaluated and validated via the *Model Evaluation Framework*), making them available for potential use within more complex applications implemented in the MATRYCS Analytics Layer. Through the gathering of streaming data from the *Streaming Module* (in the MATRYCS Governance Layer), the *Model Serving Framework* allows running the trained models in near real-time and it can thus provide the results of the ML model to the applications and services running in the MATRYCS Analytics Layer.

4.3.5.1 Design and functionality

To design the Model Serving Framework the MLOps approach has been taken in consideration like for the Model Evaluation framework. In particular MLOps comes into play to unify the two processes of developing ML models (Dev) and putting them into operation (Ops) that is the goal of this serving module. The Model Serving phase is the process that is downstream of a series of preparatory steps starting from the data acquisition following with the model training and model evaluation and needed refinement cycles. In summary, Figure 19 shows this process being managed manually, starting with the data management, moving to the model developers and their evaluation, and concluding with the model serving.

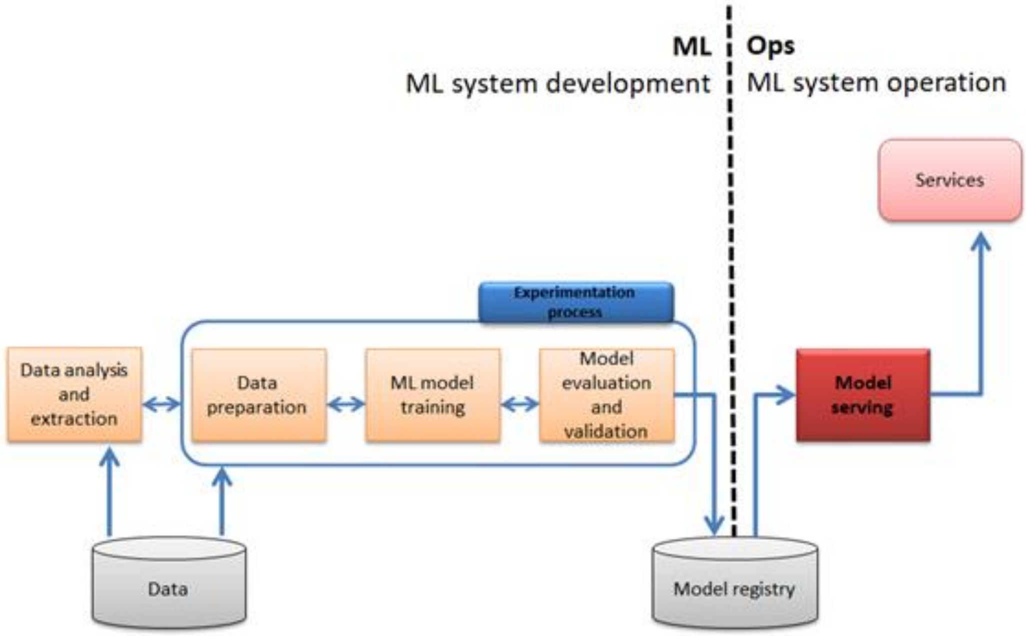


Figure 19: MLOps process – Model serving

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

To identify the best technologies and tools a series of features for the Serving framework have been taken into account. An ease to use framework also for users who are not necessarily ML experts and software engineers, the availability of APIs to be invoked directly inside the code of the project services that are going to be implemented in the context of MATRYCS project or through a graphical interface (Swagger), the presence of an easily accessible repository for loading and managing models whether external to the project or developed directly in the project, simplicity in the extension of the ML libraries adding new ones or updating the oldest in case of needs, and being ready for deployment in cloud platforms, which must be as free as possible from architectural constraints.

Below are the main characteristics of the Serving framework solution:

- The Serving Framework provides a Web UI to send prediction requests, and related APIs for model registry and deployment management. Yatai[16], model management component of BentoML, is available via Web UI, CLI, Python API.
- Production-ready online API serving and adaptive offline batch serving.
- Various ML frameworks, including Tensorflow[17], PyTorch[18], Keras[19], XGBoost[20], FastText[21], CoreML[22], Spacy[23] etc. are supported.
- Supports deployment with open-source platforms like Docker[24], Kubernetes[25], Kubeflow[26], etc. as well as one-click deployment with AWS Lambda[27], Azure Functions [28]and others, and manual cloud deployment with AWS ECS[29], Google Cloud Run[30], Heroku[31].

[16] Yatai, https://docs.bentoml.org/en/latest/api/yatai_client.html
[17] Tensorflow, https://www.tensorflow.org/
[18] Pytorch, https://pytorch.org/
[19] Keras, https://keras.io/
[20] XGBoost, https://xgboost.readthedocs.io/en/latest/
[21] FastText, https://fasttext.cc/
[22] CoreML, https://developer.apple.com/documentation/coreml
[23] Spacy, https://spacy.io/
[24] Docker, https://www.docker.com/
[25] Kubernetes, https://kubernetes.io/
[26] Kubeflow, https://www.kubeflow.org/
[27] AWS Lambda, https://en.wikipedia.org/wiki/AWS_Lambda
[28] Azure functions, https://azure.microsoft.com/en-us/services/functions/
[29] AWS ECS, https://aws.amazon.com/ecs/?nc1=h_ls&whats-new-cards.sort-by=item.additionalFields.postDateTime&whats-new-cards.sort-order=desc&ecs-blogs.sort-by=item.additionalFields.createdDate&ecs-blogs.sort-order=desc
[30] Google cloud run, https://cloud.google.com/run
[31] Heroku, https://www.heroku.com/

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

4.3.5.2 Technologies and tools

As described in previous section, purpose of the Serving Module is to serve models that have been previously trained and evaluated, and which is capable of executing ML models by having a runtime environment capable of executing as many different software libraries as required, thus allowing an almost agnostic interaction between the developer of the models and the consumer. For the heterogeneity of ML models that are going to be developed within the context of the project it is assumed that a series libraries and technologies among the most commonly used by the scientific community will be used and that the Serving Framework must be a multi-framework environment able to support models developed with technologies/libraries such as:

- Scikit-learn[32],
- Pytorch [33]
- Pytorch Lightning [34]
- Tensorflow [35]
- FastAI [36]
- Keras [37](Tensorflow 2.0 as the backend)
- fastText [38]
- CoreML [39]
- Spacy [40]
- Transformers [41]

To satisfy the identified requirement and the design choice, For all the above, the BentoML[42] framework has been adopted as the core component of the Model Serving Framework module and the aforementioned ML libraries and a dedicated repository have been installed and opportunely configured.

[32] Scikit-Learn, https://scikit-learn.org/stable/
[33] Pytorch, https://pytorch.org/
[34] Pythorch Lightening, https://www.pytorchlightning.ai/
[35] Tensorflow, https://www.tensorflow.org/
[36] FastAI, https://www.fast.ai/
[37] Keras, https://keras.io/
[38] Fasttext, https://fasttext.cc/
[39] CoreML, https://developer.apple.com/documentation/coreml
[40] Spacy, https://spacy.io/
[41] Transformers, https://huggingface.co/transformers/
[42] BentoML: https://www.bentoml.com/

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

4.3.5.3 Technical Specifications met

ID	Description of Technical Requirement satisfied	Related User Requirements
TR031	The MATRYCS platform must provide model selection.	FR005

4.4 MATRYCS Analytics Toolbox Layer

The MATRYCS Analytics Layer contains all the services that will be exposed to the end users. These services can be clustered into the Visualisations and Reports Engine, the Analytics Building Services, the Building Digital Twin and the Virtual Workbench. They are offered to the end-users via the MATRYCS Toolbox, leveraging on the different options of delivery and deployment of the cloud-based services.

4.4.1 Design and functionality

In the MATRYCS project, different analytics building services have been designed to accommodate the demands of different pilots, and to be able to cover a very wide range of possible demands from different types of users. Four main groups of services have been defined:

- MATRYCS – PERFORMANCE. It represents services able to manage all aspects related to energy performance and indoor condition of building. Specifically:
  - Energy Prediction (s1.1).
  - Building Automation Control (s1.2).
  - KPIs calculation (s1.3).
  - Technical Building Management (s1.4).
  - Optimization for network operations (s1.5).
- MATRYCS – DESIGN. It focalizes in building infrastructure. The main service is the Energy Conservation Measure (ECM)- evaluation with the related technologies catalogue:
  - Technologies catalogues services to support the design and development of building infrastructure (s2.1).
  - ECM-based scenarios evaluation service (s2.2).
- MATRYCS POLICY that provides services for policy making and policy impact assessment. The tools are:
  - SECAPs decision-making support (s3.1).
  - Energy Performance Certification harmonization and checking (s3.2).
  - National and EU policy impacts assessment and support (s3.3).

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

- MATRYCS FUND in which analytics for de-risking investments in energy efficiency will be developed. The tools are:
  - Services to support Energy Savings M&V towards improved Energy Performance Contracts (s4.1).
  - Services to support the financing of the EE refurbishments (s4.2).

Together with these groups, two other tools will be developed to cover fundamental aspects of building design and management. The two tools are Building Digital Twins (s0.1) [11] and the geoclustering tool (s0.2).

All tools will be stand-alone software, that can be installed both on-premises and on cloud. A key point for the functioning of the services is the access to the data. The latter will be carried out in three ways:

- direct access to building database
- connection to specific frameworks where different machine learning algorithms can run generating data and analysis to be shown in a dedicated front-end.
- Using the reasoning engine.

For proper operation, the services leverage most of the components developed within MATRYCS GOVERNANCE(WP3) and MATRYCS PROCESSING (WP4). The General Reference Architecture (Figure 5) a general overview of the connection of all analytical services with the MATRYCS infrastructure is provided, including specific examples for one analytical service (Figure 21). For more details on the development of individual services and the resources used, please refer to the deliverables of WP5 (D5.1, D5.2 and D.3). At this stage only deliverable D5.1 is present, the remaining deliverables will be presented at month 24 (September 2022) and at month 31 (April 2023). They represent a constant update of the development of the analytical services from the initial version (D5.1) to the deployed version (D5.3).

As it possible to see in the Figure 21, Depending on the type of service, determining components of WP3 and WP4 are used. In this specific case (geoclustering tool - s0.2), the tool uses the WP3 reasoning engine service to obtain the necessary data to be used both as front-end input for simple visualization and for the generation of Machine Learning models (Clustering algorithm) to be provided to the serving framework. The whole connection is made through the use of specific APIs that are developed on top of each component.

The services are developed with different technologies but are deployed in the Matrycs infrastructure as docker containers and separated as IaaS, PaaS and SaaS [12] . Each container works independently and manages all connections to the different components of Matrycs governance and processing according to the usage required for proper operation. Using the EGI infrastructure the dockers are deployed within virtual machines each of which has the resources required for the docker to work (Figure 20).

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

The analytical services connect directly to WP3 and WP4 which are also deployed within the EGI infrastructure. The communication between the different components is through API or restAPI depending on the type of connection with the different components.

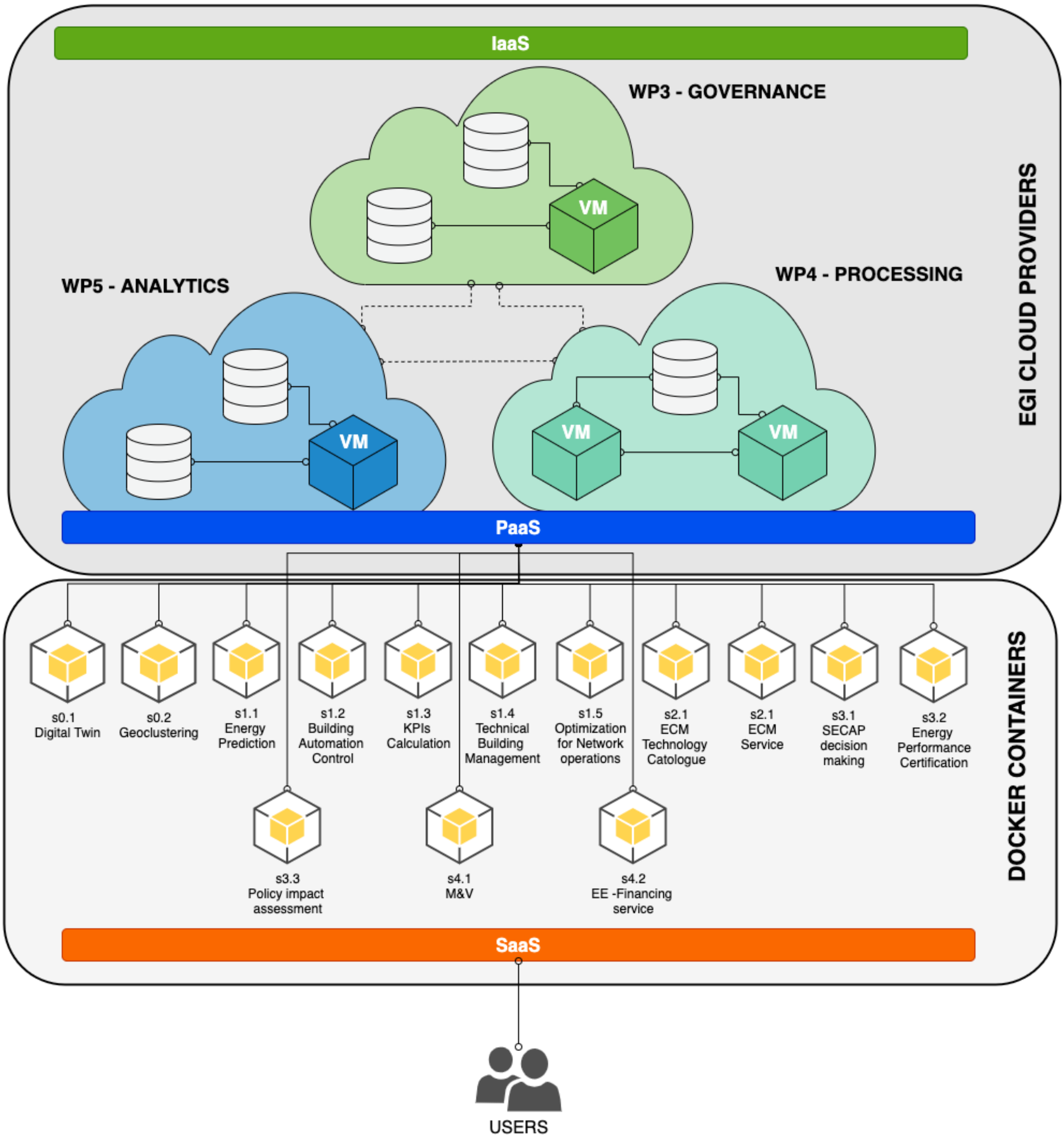


Figure 20 Analytics services on EGI infrastucture

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

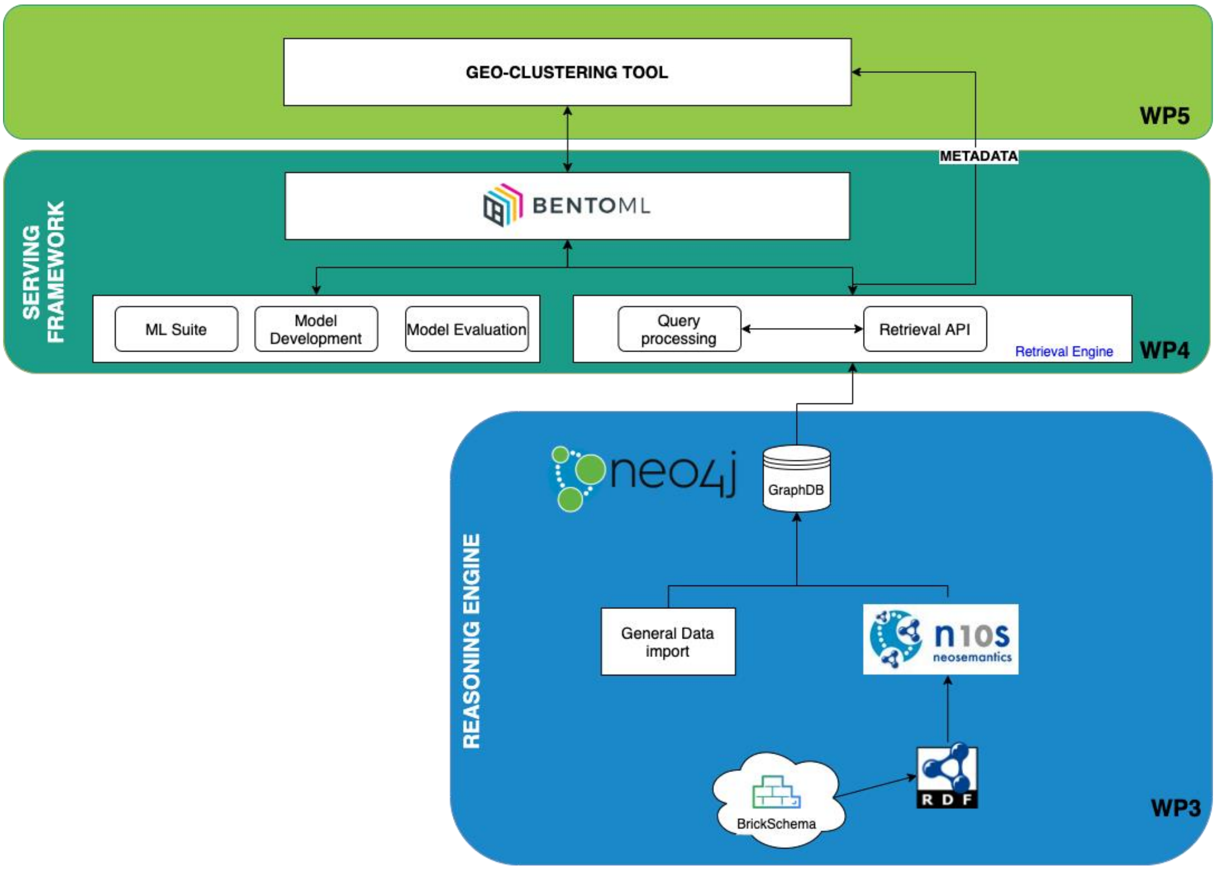


Figure 21 Geoclustering tool (s0.2) connection with components of WP3 and WP4

4.4.2 Technologies and tools

All services are developed as web applications. Different technologies have been implemented depending on the different service.

In general, almost all applications are written using python (flask or django) in the backend while the frontend uses different technologies (react.js, next.js, angular, etc.). Each service is developed to run independently from others and provided as a docker container in the deployment phase. This choice allows on the one hand the easy management of services by each developer and the other to simplify both the resources needed for deployment and maintenance.

A different path is taken for the digital twin service where specific technologies related to BIM are taken into account in order to generate building models able to provide useful information to the user through the association of data from sensors. In this case, open ifc visualizers combined into web frameworks are used.

4.4.3 Technical Specifications met

Table 22 MATRYCS Analytics Toolbox Layer Technical Specifications

ID	Description of Technical Requirement satisfied	Related User Requirements

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

TR020	The MATRYCS platform must provide alarm messages.	FR052
TR026	The MATRYCS platform must provide user-friendly templates for variable specification.	FR026
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, FR009, FR015, FR017, FR018, FR019, FR025
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
TR058	The MATRYCS platform must allow service selection by the user.	FR022
TR059	The MATRYCS platform must provide the development of buildings' digital twin.	FR044
TR060	The MATRYCS platform must provide a digital twin service with interoperability functionalities.	FR037
TR061	The MATRYCS platform should provide a functionality of new project/activity creation.	FR046

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

4.4.4 Virtual Workbench

4.4.4.1 Design and functionality

The Virtual Workbench provides SMEs, developers, engineers, data scientists, researchers, and potential innovators a graphical UI to develop in scale new services, write algorithms, define applications, develop micro-service and functions taking advantage from the available Matrycs assets such as datasets and ML models. The possibility to have both data and ML models easily accessible under a single environment facilitates the creation and testing of potential new services at scale and reduces the time required for developers and SMEs to create them.

The module is mainly intended for a user of experts able to write, run, test, update and save their created functions under a unique environment ready to run the algorithms and show elaboration results to be analysed.

Two main sub-modules are included in the MATRYCS Virtual Workbench:

- The *Building Data and Services Innovation Hub*;
- The *Data Analytics Environment*.

The *Building Data and Services Innovation Hub* is the module aimed at providing the functionalities to facilitate the creation of services quickly and easily through an intuitive user interface. The available datasets and the pre-trained models exposed by the MATRYCS Processing Layer through the *Model Serving Framework* are made available in the Virtual Workbench for analysis and reuse.

The *Data Analytics Environment* is an additional component that allows users (e.g., analysts, data scientist) to perform freeform queries and data analytics on the assets that are accessible to the platform (models and data).

4.4.4.2 Technologies and tools

As from GA, the Knowage[43] suite developed by Engineering II is the platform identified to provide the Virtual Workbench functionalities, it is open-source with the community and enterprise-level solutions. The version adopted for the Matrycs project is the community one that is constantly updated with the new functinalities developed for the the project and related update during the testing and usage phase. The suite has been installed using Docker technology, in particular four Docker container made the entire suite and each of these is designed and developed to run the suite services as unique application being interconnected. Detail about the software component cab be found in the official Knowage website and github repository[44].

[43] Knowage suite: https://www.knowage-suite.com/site/
[44] Knowage GitHUB: https://github.com/KnowageLabs

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

4.4.4.3 Technical Specifications met

Table 23 Virtual Workbench Technical Specifications

ID	Description of Technical Requirement satisfied	Related User Requirements
TR014	The MATRYCS platform must allow data connection with MATRYCS Governance layer.	FR035, FR039, NFR010
TR015	The MATRYCS platform must allow the user to upload datasets and ML/AI models.	FR027
TR019	The MATRYCS platform must show false usage messages when misusage of the platform occurs.	FR053, NFR022, NFR027
TR024	The MATRYCS platform must enable verification of the identity of all users using the platform.	FR033
TR028	The MATRYCS platform must provide user navigation and functionalities of the platform in a convenient way.	FR034, FR047
TR030	he MATRYCS platform must support search functionality over the datasets and ML models.	FR049, FR051, FR056
TR039	The MATRYCS platform's usage must be supported by user guidelines, concerning both navigation and data uploading.	NFR025, NFR026
TR040	The MATRYCS platform must offer easy user navigation through the platform.	FR038, NFR024

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

4.4.5 Advanced Visualisations & Reports Engine

4.4.5.1 Design and functionality

Visualisations & Reports Engine has been designed and implemented under Task 4.4 – Advanced Visualisations and Reports Engine and it is a user-friendly front-end application, which offers a variety of capabilities. Users can create several different types of charts (e.g., pie charts, bar charts, gauge charts, line charts) over the stored MATRYCS data, as well as create responsive tables and map visualisations as far as the data include co-ordinate information, that display the requested data. Requests to the Visualisation & Reports are created by either filling properly designed forms, or by inserting plain SQL code, which is formatted in the background and sent to the backend via REST-API requests. Additionally, user profile page is enabled, which displays the user's basic information, user's roles, latest queries performed and files that have been uploaded. Uploaded files are also available for querying, but only for the respective user that uploaded them.

Visualisations & Reports Engine is fully integrated with the End-to-End Security Framework and the access role management is implemented within the component. Specifically, non-signed-in users cannot have access to any visualisation and reporting tools. After signing-in, the users can navigate to the engine's pages, as well as their personal profile page, which displays the information as described above. Moreover, by leveraging the Security Enabler's functionalities, (developed under task T3.6 – End-to-End Security Framework), it is possible to limit the access to specific datasets for specific users, if their roles do not include the ones that can access them.

Figure 22 presents a high-level architecture and usage flow of the Visualisations & Reports Engine. The technologies used are indicated and will be further explained in Section 4.3.4.2 – Technologies and Tools.

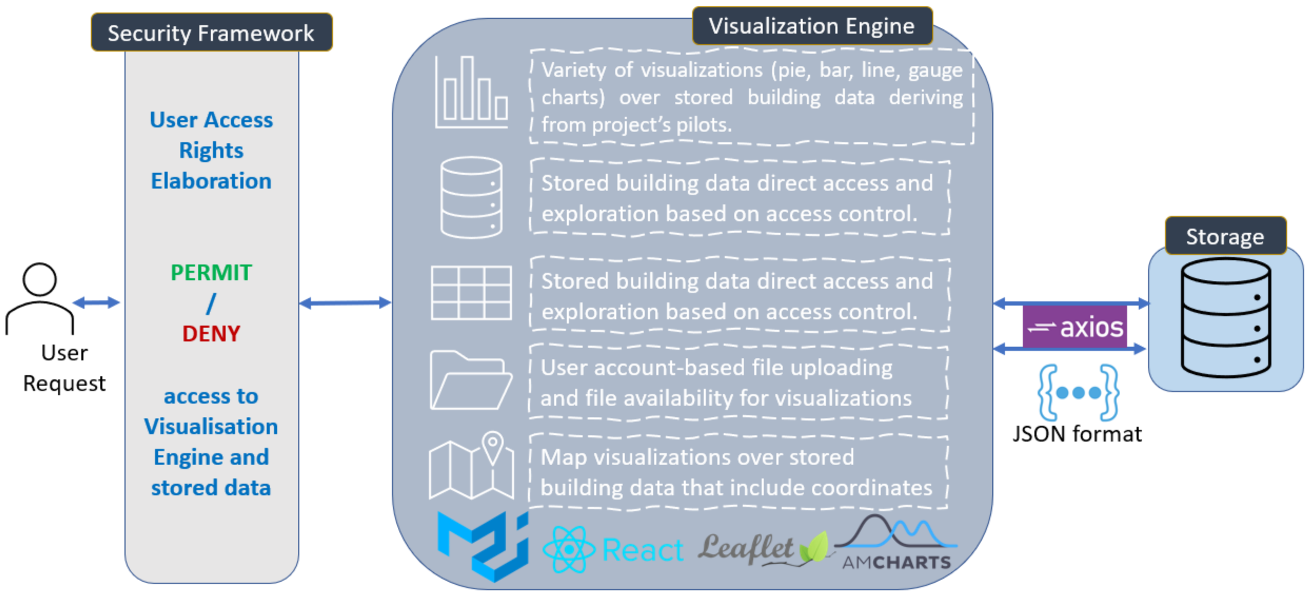


Figure 22: Visualisation Engine and connection with Security Framework high-level architecture

4.4.5.2 Technologies and tools

The technologies used for the creation of the Visualisations & Reports Engine, derived after a

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

thorough state-of-the-art analysis; a process in which the best suiting technologies were identified, libraries, and tools that would assist in the creation of a robust and reusable tool.

These technologies are listed, along with a short description, in Table 24.

Table 24 Tools & Technologies used in Advanced Visualisations & Reports Engine

Technologies / Tools	Description
JavaScript	Programming Language
React	Free, open-source, front-end JavaScript library that is user for building single-page applications / user interfaces based on components.
MaterialUI	A set of React-based components which are simple and customizable.
AmCharts	A set of JavaScript-based data visualisation libraries that includes regular chart types like Pie Charts and Bar Charts, as well as advanced ones like Stock Charts and Maps.
Leaflet	An open-source JavaScript library used to build web mapping applications.
Axios	A JavaScript library used to make HTTP requests.

4.4.5.3 Technical Specifications met

Table 25 Advanced Visualisations & Reports Engine

ID	Description of Technical Requirement satisfied	Related User Requirements
TR034	The MATRYCS platform must provide data visualisation tools and functionalities (graphs, maps etc.).	FR070, FR071, FR073, FR074, NFR023
TR035	The MATRYCS platform must provide reporting tools and services to report the generated numerical results.	FR057, FR060, FR061, FR062, FR063, FR065
TR036	The MATRYCS platform must provide services to propose list of actions.	FR055, FR058, FR059, FR064, FR066, FR067

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

4.5 End-to-end security

The end-to-end security in the MATRYCS Reference Architecture is implemented as an end-to-end security framework based on four pillars. The framework provides a set of design decisions and implementation aspects related to high-level security, fine-grained access control and privacy, encompassing various MATRYCS entities – also associated with big data for buildings: infrastructure, assets, services, end-users, and data. The end-to-end security framework supports inter-entity secure encrypted communication and enables mechanisms for maintaining and enforcing appropriate legal, security and privacy policies, thus increasing the system's trustfulness. Additionally, it ensures anonymization and stored data encryption where required. For auditing purposes, a logging system is employed. Moreover, the end-to-end security framework provides means for a system-wide authentication and authorization as well as software vulnerabilities/flaws detection and mitigation tools. More details about the end-to-end security framework can be found in the MATRYCS Deliverable D3.2 (released in M11, i.e., August 2021).

4.5.1 Technical Specifications met

Table 26 End-to-end Security Technical Specifications

ID	Description of Technical Requirement satisfied	Related User Requirements
TR023	The MATRYCS platform must ensure different authorisation levels for accessing datasets.	FR033, NFR016, NFR019
TR025	The MATRYCS platform must provide a secure and controlled registration process for new users	NFR016, NFR018
TR038	The MATRYCS platform must forbid unauthorised user access to the platform and the datasets.	FR033, NFR019

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

5 Mapping of MATRYCS Reference Architecture to existing architecture models

This Section aims at showing how the MATRYCS Reference Architecture relates to the other main architecture solutions presented in the context of big data (see Section 2). The goal is to determine if the different architectural views are consistent and, in case, to identify possible existing gaps in the architecture definitions. The alignment of the MATRYCS Reference Architecture with respect to the BDVA Reference Model, to the AIOTI HLA, to the FIWARE Open Reference Architecture has been descripted in D2.3.

The following of this Section presents the alignment of the MATRYCS Reference Architecture with respect to the OpenDEI, to Bridge, and to IDS.

5.1 Alignment with OpenDEI Reference Architecture

Similar like the BDVA Reference Model, the OpenDEI [14] does not provide a strict definition of architectural layers, but rather it presents the horizontal and vertical concerns that are relevant within a big data reference architecture. The MATRYCS Reference Architecture maps very well to the OpenDEI Reference Model, namely it covers the most existing horizontal and vertical concerns. Figure 23 shows the mapping between the two architecture views.

Looking at the horizontal concerns, the bottom layer of the MATRYCS Reference Architecture includes all the possible data sources, thus including sensors, actuators, IoT devices, but also other platforms and open datasets. The MATRYCS Governance Layer covers three of the BDVA horizontal concerns. The MATRYCS *Interoperability Module* and *Data Pre-processing Module* cover in fact the concerns within the BDVA Data Management and specific functions are foreseen in the *Data Pre-processing Module* to guarantee data protection. The management of both data at-rest and data-in-motion required in the BDVA Data Processing block is instead guaranteed in the MATRYCS architecture via the *Data Streaming Module* and the *Query Engines* associated to the *Data Storage*s. The suite of ML/DL trained models in the MATRYCS Processing Layer and the *Analytics & Building Services* in the MATRYCS Analytics Layer provide the analytics capabilities highlighted in the BDVA Data Analytics, whereas the *Visualization and Reporting Engine* of the MATRYCS Analytics Layer allows visualization and user interaction.

For the vertical concerns, security is of course one the major challenges to be addressed in both architectures. Regarding the concern of DATA SHARING, the MATRYCS Reference Architecture supports all the three functions in OpenDEI: Transact. Manager (DLT), Data Models/Ontologies and Data Sharing API. This is illustrated as dash line in from all the three layers, which means the entire layer is able to share data through those three functions. The MATRYCS Processing layer provide the functionality to analysis the data with different AI services, which help to enable effective data trading and business management. The last part of the vertical concern is the digital transformation x-industry pilots, which is the different services (monitoring, maintenance, quality etc.) can be provided to the pilot according to the OpenDEI. The MATRYCS Architecture offers the MARTRYCS Toolbox, which contains four categories of services: funding, design, performance and policy. Because MATRYCS is more interesting in the building domain applications. Other services are also able to implement and add to MATRYCS Architecture.

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

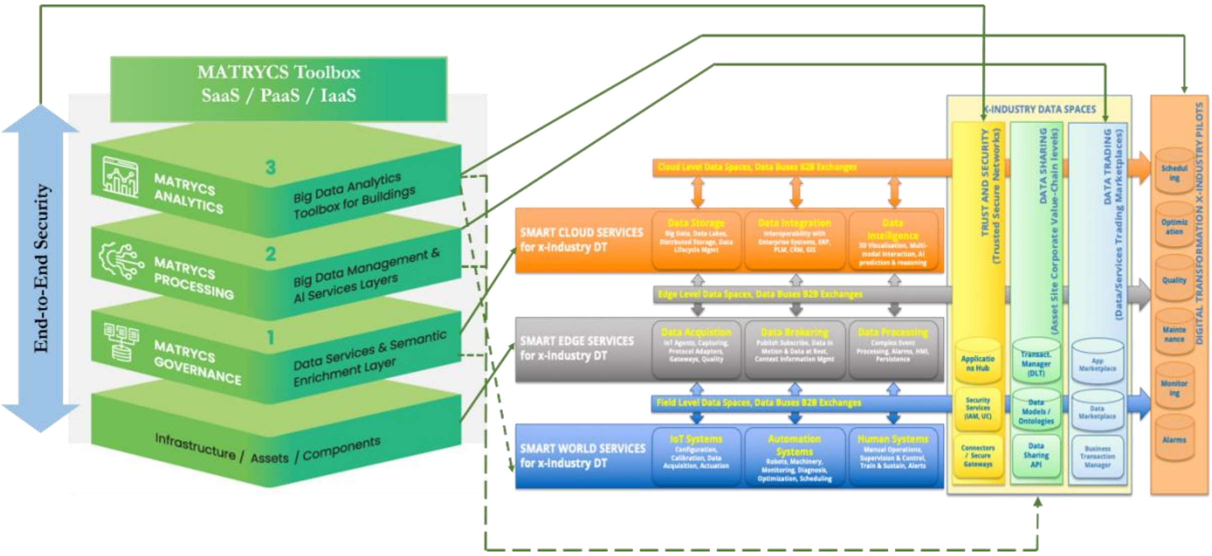


Figure 23 Mapping of MATRYCS Reference Architecture to OpenDEI Reference Architecture

5.2 Alignment with BRIDGE Reference Architecture

BRIDGE proposes an Open Reference Architecture that has many aspects in common with the MATRYCS Reference Architecture, especial because it has the same goal but only in different domain. As a consequence, not only the mapping of the two architecture sounds straightforward, but also the MATRYCS Reference Architecture have used the same standard and component. Figure 24 shows the mapping between the two architectures.

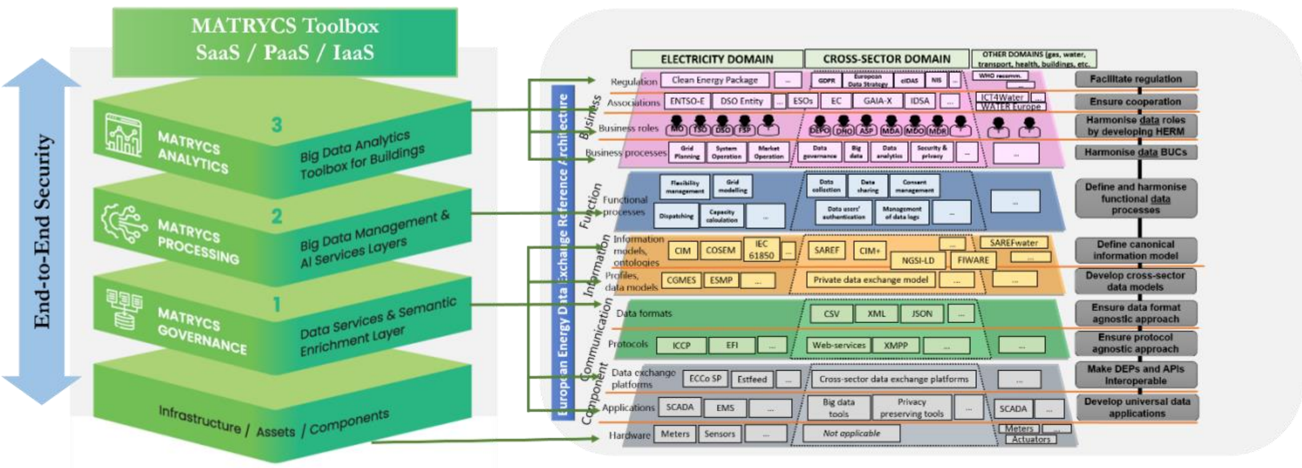


Figure 24 Mapping of MATRYCS Reference Architecture to BRIDGE Reference Architecture

BRIGE is define according to five layers: Component, Communication, Information, Function and Business.

Looking at the Component layer, the bottom layer of the MATRYCS Reference Architecture includes all the possible data sources, thus including sensors, actuators, IoT devices, but also other platforms and open datasets. The Communication layer and Information layer is covered by the MATRYCS Governance Layer, which contains the MATRYCS *Interoperability Module* and *Data Pre-processing*

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

Module to cover in fact the concerns with protocols. The MATRYCS *Common Data Module* and *Semantic Enrichment Model* have the same functionality with BRIGE information models and data models. There are many same standard, data model and ontology between those two parts like SAREF, SAREF extension, FIWARE and NGSI-LD.

The suite of ML/DL trained models in the MATRYCS Processing Layer and the *Analytics & Building Services* in the MATRYCS Analytics Layer provide the analytics capabilities highlighted in the Function layer of BRIDG, whereas the *Visualization and Reporting Engine* of the MATRYCS Analytics Layer allows visualization and user interaction and makes the data management and data usage more easily.

Finally, regarding the concern of Business layer, it is worth noting that the MATRYCS architecture offers the MATRYCS Toolbox, where different business users, engineers, developers and associations can all have access at different functionalities provided as SaaS, PaaS and IaaS. The business processes in BRIGE contains the aspect about big data, data governance, data analytics and security and privacy, which is the same aspect in MATRYCS as mentioned before in Section 4.

At the top right of BRIGE RA, there are OTHER DOMAINS, where MATRYCS RA can be integrated as buildings domain perfectly according to the Alignment that is descripted before.

5.3 Alignment with Data Space

Data Space is a decentralized data ecosystem based on universally agreed-upon building pieces that allows members to share data effectively and safely. There are three technology building blocks: Data interoperability, Data Sovereignty and trust and Data value creation, which are illustrated in Figure 25 from [11]. [11] from FIWARE for Data Spaces defines the complete taxonomy of building blocks required for creating data spaces.

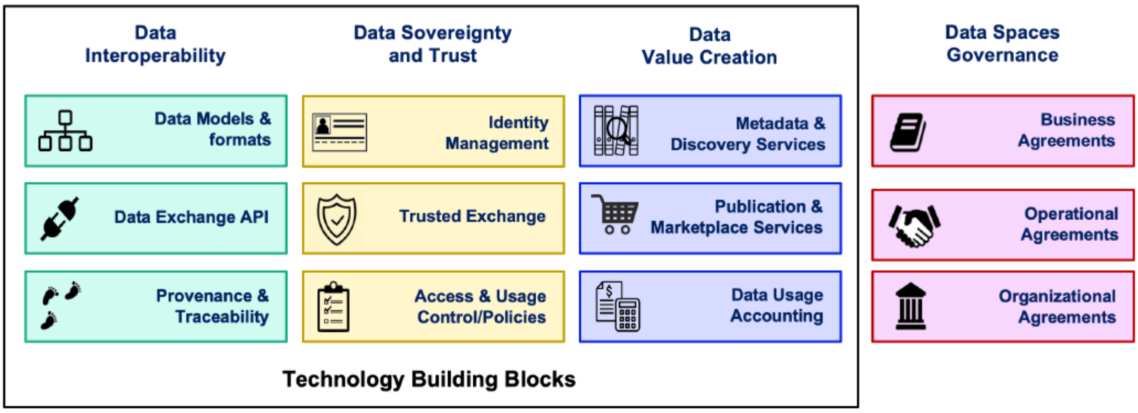


Figure 25 Building blocks in a Data Space

As described before, IDSA provides various architecture aspects that are required to establish data spaces, the most essential of which are trust and data sovereignty. These are important but insufficient for the creation of actual live data spaces that are recognized and modified in the marketplace. FIWARE have the strength reside in Standard APIs, data models, and marketplace features. Figure 26 shows how the two efforts complement one other in terms of technical building pieces. The presence of the IDSA emblem on a building block indicates that requirements have been developed or are

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

being developed by IDSA. The FIWARE logo indicates that the FIWARE Community has developed open-source component implementations and, eventually, is driving them.

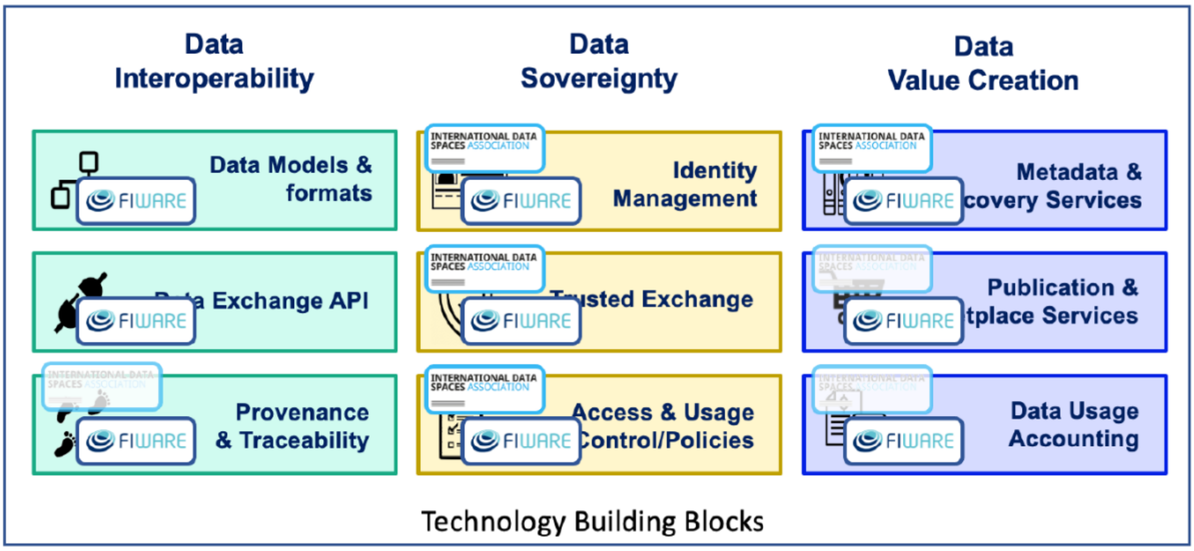


Figure 26 IDSA and FIWARE positioning to create Data Spaces

Based on the Figure 26, the Figure 27 is created and shows the mapping of MATRYCS with FIWARE and IDSA. Starting from the MATRYCS Governance Layer, this has the same functionality with the Data Interoperability. It is the MARTRYCS Common Data Model, APIs and Query Engine. The End-to-End Security framework in MATRYCS is based on the FIWARE, which leads to direct mapping between each other. Therefore, the Data Sovereignty is mapped with MATRYCS, FIWARE and IDSA. The Last part is Data Value Creation, which contains Metadata & Discovery Services, Publication & Marketplace and Data Usage Accounting. In MATRYCS Processing layer, different AI services are developed, which will analysis the data and try to add value to the data. However, the mapping is not straightforward. Therefore, the dash line is used. The MATRYCS Analytics toolbox is also mapped to the Data Value Creation part. Because MATRYCS comes with a marketplace (and the associated functionalities for buying and selling data) as well as with the possibility of adding and developing tools.

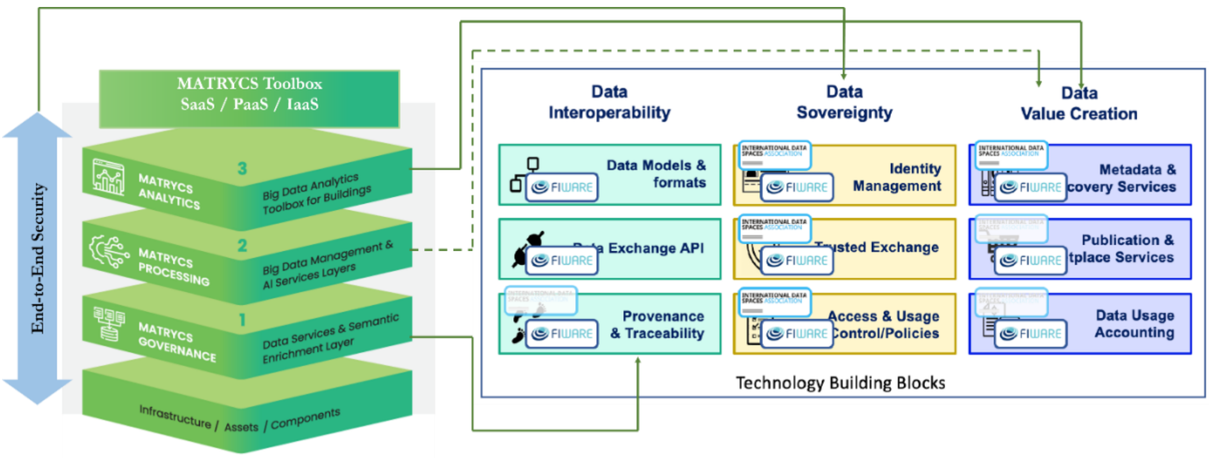


Figure 27 Mapping of MATRYCS Reference Architecture IDS and FIWARE Reference Architecture

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

6 Conclusion

In this Deliverable, a final version of the MATRYCS Reference Architecture for the building domain has been presented, which is based on the first version of the MATRYCS Reference Architecture. The proposed architecture takes into account already existing concepts and developments in the area of Big Data Architectures as well as the specifications and requirements derived from the analysis of the updated version of the User Stories, Usage Scenarios and Use Cases is presented to provide a high-level, but at the same time comprehensive, description of the proposed Reference Architecture for building data spaces.

The final version of the MATRYCS Reference Architecture provide two viewpoints of the architecture model, which are Technology View and Functional View. Compared with the first version, which only provide a short overview of the core components, the technology detail of each module and the technical specification met.

This Deliverable added the analysis of the mapping of the three other well-known Reference Architecture Frameworks. This allows identifying the connections among different architecture models as well as the possible gaps to be considered for refining the design of the MATRYCS architecture.

It is worth noting that, as this Deliverable concerns the final version of the MATRYCS Reference Architecture, the proposed concepts, definitions and models are updated based on the discussion and partially modified over the course of the project activities. In this regard, future steps compared to the first version include the extension of the survey on existing Reference Architecture specifications as well as the review of the specifications and requirements derived from the MATRYCS use cases, following the further developments in each one of the Large-Scale Pilots has been done.

A refined description of the specifications of each architecture component are provided according to the progress made in the project. Finally, the alignment of the MATRYCS Reference Architecture with other architecture descriptions (with the BRIGE, OpenDEI and IDS Reference Architecture Framework) conceived in similar projects in the domain of building platforms and big data are conducted and detailed.

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

References

[1] GAIA-X association "GAIA-X: Technical Architecture" , [online] https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-technical-architecture.pdf?__blob=publicationFile&v=5

[2] Data Management Working Group, "BRIDGE: European energy data exchange reference architecture",[online] https://ec.europa.eu/energy/sites/default/files/documents/bridge_wg_data_management_eu_reference_architcture_report_2020-2021.pdf

[3] MATRYCS Project, "D2.1 State-of-the-art analysis and Big Data Value Chain", Apr. 2021, [online] https://matrycs.eu/resources/results

[4] MATRYCS Project, "D2.2 - MATRYCS Technical and Security Specifications", Apr. 2021, [online] https://matrycs.eu/resources/results

[5] MATRYCS Project, "D2.3 - MATRYCS Reference Architecture for Buildings Data v1.0", Apr. 2021, [online] https://matrycs.eu/resources/results

[6] MATRYCS Project, "D3.1 – MATRYCS-GOVERNACE (1st technology release)", Apr. 2021, [online] https://matrycs.eu/resources/results

[7] MATRYCS Project, "D4.1 – MATRYCS-PROCESSING (1st technology release)", Apr. 2021, [online] https://matrycs.eu/resources/results

[8] MATRYCS Project, "D4.4 – MATRYCS-ANALYTICS and Open Cloud-based Data Analytics Toolbox (1st technology release)", Apr. 2021, [online] https://matrycs.eu/resources/results

[9] MATRYCS Project, "D5.1 – MATRYCS Analytics Building Services V1.0", Apr. 2021, [online] https://matrycs.eu/resources/results

[10] Industrial Internet Consortium, "The Industrial Internet of Things Volume G1: Reference Architecture", 2017, [online] http://www.iiconsortium.org/IIRA.htm

[11] Shahzad, M.; Shafiq, M.T.; Douglas, D.; Kassem, M. Digital Twins in Built Environments: An Investigation of the Characteristics, Applications, and Challenges. Buildings 2022, 12.

[12] Sowmya, S.; Deepika, P.; Naren, J. Layers of cloud – IaaS, PaaS and SaaS: a Survey. International Journal of Computer Science and Information Technologies 2014, 5

[13] FIWARE Foundation, "FIWARE for Data Spaces – version 1.0", https://www.hannovermesse.de/apollo/hannover_messe_2021/obs/Binary/A1085838/FIWARE%20for%20Data%20Spaces%20%281%29.pdf

[14] OPEN DEI Project, "D2.1 – Reference Architecture for Cross-Domain Digital Transformation V1", Nov. 2020, [online] https://www.opendei.eu/case-studies/d2-1-reference-architecture-for-cross-domain-digital-transformation/

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

Annex I: User stories, usage scenarios, use cases

This annex presents the complete user stories behind all the use cases considered in the MATRYCS platform so far. These are presented separately in LSP tables.

Table 27: LSP 1 Usage scenarios and user stories considered in MATRYCS

LSP 1 – Usage scenarios and User stories considered in MATRYCS			
UC	#USC	USC	User stories
UC01_01	4	USC 1.2.1	**US 1.2** As an Informed facility manager working for the large shopping, entertainment, business, commercial and logistics centre, I want to get reliable information to do preventive maintenance of the most critical systems so that I can implement optimal strategy to improve overall reliability and efficiency of our technical systems.
		USC 2.2.1	**US 2.2** As a Sustainability supporting facility manager, I want to objectively evaluate all positive and potentially negative effects of executed preventive maintenance measures so that I can be aware of sustainability aspects of all planned and implemented measures and make more reliable implementation plans for the context of each building.
		USC 3.1.1	**US 3.1** As a Cost-conscious facility manager working for the large shopping, entertainment, business, commercial and logistics centre, I want to be capable to do the first pre-feasibility evaluations of all identified performance improvements so that I can provide the reliable evaluation of the available energy and maintenance improvement opportunities for the context of each building and provide sound decision support for the management board.
		USC 3.2.1	**US 3.2** As a Cost-conscious facility manager working for the large shopping, entertainment, business, commercial and logistics centre, I want to create reliable implementation action plans so that I can more closely monitor their implementation and propose eventual corrective measure to properly mitigate risks related with potential deviations.
UC01_02	7	USC 1.1.1	**US 1.1** As an Informed facility manager working for the large shopping, entertainment, business, commercial and logistics centre, I want to create reliable implementation action plans for identified energy efficiency and facility management measures so that I can propose optimal solutions to implemented in each building and provide sound decision support for the management board.
		USC 2.1.1	**US 2.1** As a Sustainability supporting facility manager, I want to create reliable monitoring and evaluation plans and procedures for selected energy efficiency, management and maintenance measures so that I can be aware of sustainability aspects of all planned and implemented measures and make more reliable implementation plans for the context of each building.
		USC 5.1.1	**US 5.1** As a Sustainability supporting owner of shops and offices at the large shopping, entertainment, business, commercial and logistics centre, I want to become aware of the sustainability aspects of the proposed energy and facility measures so that I can implement optimal solutions and improve overall sustainability of my business.
		USC 5.3.1	**US 5.3** As a Sustainability supporting owner of shops and offices at the large shopping, entertainment, business, commercial and logistics centre, I want to create greener image of my property (offices and shops) so that I can reach more demanding clients and improve competitiveness of my business.
		USC 1.4.1	**US 1.4** As an Informed facility manager working for the large shopping, entertainment, business, commercial and logistics centre, I want to be empowered to validate achieved energy and cost savings so that I can independently prepare necessary reports and help my company to report about achieved savings, implement potential corrective measures and improve overall image of the company.

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

		USC 4.1.1	**US 4.1** As an Informed owner of shops and offices at the large shopping, entertainment, business, commercial and logistics centre, I want to become aware of energy and comfort improvement opportunities for my property so that I can implement the most effective measures to reduce energy consumption and related costs and at the same time improve the user comfort for my workers and clients.
		USC 5.2.1	**US 5.2** As a Sustainability supporting owner of shops and offices at the large shopping, entertainment, business, commercial and logistics centre, I want to become aware of energy and environmental footprint of my property so that I can implement optimal solutions to improve overall sustainability of my business and the same time improve the user comfort for my workers and clients.

Table 28: LSP 2 Usage scenarios and user stories considered in MATRYCS

LSP 2 – Usage scenarios and User stories considered in MATRYCS			
UC	#USC	USC	User stories
UC02_01	2	USC 2.1.1	**US 2.1** As a worker of the City of Gdynia (owner of the educational buildings), I want to have access to all the information about building that are available (e.g. geometrical information, energy analysis) so that I can plan well the renovation activities. I want to know the best retrofitting solutions for a certain building and energy predictions, so that I have a reliable information in the context of each building.
		USC 3.1.1	**US 3.1** As a designer/architect that is performing the renovation design, I want to know which renovation scenario fits the best to the building needs so that I can make a good proposition for the renovation for the investor.
UC02_02	2	USC 1.1.1	**US 1.1** As a worker of the kindergarten, I want to have access to the information about indoor air quality and energy consumption so that I can increase the comfort in my working place.
		USC 3.1.1	**US 2.2** As a worker of the City of Gdynia (owner of kindergarten no 16 in Gdynia), I want to know to what is the indoor air quality factor inside the buildings. This factor should be calculated based on the installed monitoring devices (monitoring of humidity, temperature, co2). Also, interesting could be what further intervention on the building can be done so it meets NZEB standard.

Table 29: LSP 3 Usage scenarios and user stories considered in MATRYCS

LSP 3 – Usage scenarios and User stories considered in MATRYCS			
UC	#USC	USC	User stories
UC03_01	3	USC 1.1.1	**US 1.1** As an Advanced ESCO, I want to implement energy efficiency measures in different facilities, so that the efficiency of these facilities decreases, reducing costs and GHG emissions.
		USC 1.4.1	**US 1.4** As an Advanced ESCO, I want to make the energy savings verification procedure easier, so that the increase of the owners can increase.
		USC 2.3.1	**US 2.3** As an EPC facilitator, I want to provide accurate information about the facilitie's behaviour before and after and intervention so that owners realise the benefits provided by energy efficiency measures.
UC03_02	3	USC 1.1.1 USC 1.4.1 USC 2.3.1	Same three user stories as above.

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

Table 30: LSP 4 Usage scenarios and user stories considered in MATRYCS

LSP 4 – Usage scenarios and User stories considered in MATRYCS			
UC	#USC	USC	User stories
UC04_01	3	USC 7.1.1	**US 7.1** As a Small-Scale DSO Head, I want to invest in a new system of monitoring and control so that I have an investment plan to implement the system.
		USC 7.2.1	**US 7.2** As a Small-Scale DSO Head, I want to improve the system stability so that I can exploit flexibility from the buildings and assets of my distribution network.
		USC 7.3.1	**US 7.3** As a Small-Scale DSO Head, I want to minimize the congestions in a district related to the increase of EVs so that I have a charging session rescheduling infrastructure to manage the power flows.
UC04_02	3	USC 1.1.1 USC 1.1.2	**US 1.1** As a Sensitive office employee working at the company headquarter, I want to improve the air conditioning and lighting systems so that I have better workplace conditions through a new innovative system.
		USC 1.2.1	**US 1.2** As a Sensitive office employee, I want to create a comfortable work environment derived through the green energy so that I contribute to the energy transition.

Table 31: LSP 5 Usage scenarios and user stories considered in MATRYCS

LSP 5 – Usage scenarios and User stories considered in MATRYCS			
UC	#USC	USC	User stories
UC05_01	3	USC 2.1.1	**US 2.1** As an EV owner who are willing to reduce my carbon footprint, I want to be able to know the carbon footprint of the electricity flowing in my EV so that I can truly quantify the carbon footprint reduction of my EV.
		USC 2.2.2	**US 2.2** As an EV owner who are willing to reduce my carbon footprint, I want to know in advance when the electricity coming from the grid is "greener" so that plan my EV charges to further reduce my carbon footprint
		USC 5.1.1	**US 5.1** As a eco-friendly SME who would like to further increase my positive impact on the local community, I want to learn more about my energy consumption so that I can understand how can I further reduce my overall consumption and carbon footprint linked to it.
UC05_02	2	USC 1.1.1 USC 1.1.2	**US 1.1** As a Prosumer that has decided to invest part of his/her savings to install a PV system, I want to be able to identify when my PV system is not performing properly so that I can minimise the "downtime" and maximize the profit of my investment
UC05_03	3	USC 1.2.1 USC 1.2.2 USC 1.2.3	**US 1.2** As a Prosumer that has decided to invest part of his/her savings to install a PV system, I want to be able to identify when my PV system is not performing properly so that I can minimise the "downtime" and maximize the profit of my investment
	1	USC 3.2.1	**US 3.1** As an environmentally conscious citizens who doesn't have significant investment power, I want better understand my electricity (and energy) consumptions so that I learn how can I reduce my electricity bills.
UC05_04	1	USC 3.2.1	**US 3.2** As an environmentally conscious citizens who doesn't have significant investment power, I want to assess the environmental and economic consequences of energy efficiency investment so that I can invest wisely my savings.

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

Table 32: LSP 6 Usage scenarios and user stories considered in MATRYCS

LSP 6 – Usage scenarios and User stories considered in MATRYCS			
UC	#USC	USC	User stories
UC06_01	1	USC 6.1.1	**US 6.1** As an Informed owner, I want to be able to prove the benefits of the renovation of the District Heating Network, so that I can convince other owners to carry of the renovation.
UC06_02	1	USC 6.1.1	Same user story as above.
UC06_03	7	USC 1.1.1	**US 1.1** As a Sustainability supporter DHN manager, I want to increase my knowledge about the operation of the facility, so that the O&M activities are easier and cheaper to carry out.
		USC 1.2.1	**US 1.2** As a Sustainability supporter DHN manager, I want to predict accurately the behaviour of the facility and its response to possible changes, so that there is a reduction on the energy consumption and the GHG emissions of the facility.
		USC 1.3.1	**US 1.3** As a Sustainability supporter DHN manager, I want to have an accurate model of the District Heating Network, so that it is easier to expand the network and to implement new energy efficiency measures.
		USC 2.1.1	**US 2.1** As a Cost-conscious facility manager, I want to increase my knowledge about the operation of the facility, so that the efficiency of the facility is higher and the costs are reduced.
		USC 2.2.1	**US 2.2** As a Cost-conscious facility manager, I want to predict accurately the behaviour of the facility and its response to possible changes so that the risks of implementing energy efficiency measures are reduced
		USC 2.3.1	**US 2.3** As a Cost-conscious facility manager, I want to have an accurate model of the District Heating Network, so that the risks of expanding the network and its costs are reduced.
		USC 4.3.1	**US 4.3** As an Efficiency-oriented designer, I want to allow the manager to have an accurate model of the network so that he is able to manage it properly.
UC06_04	5	USC 3.1.1	**US 3.1** As a Profit-oriented DHN designer, I want to increase my knowledge about DHN operation, so that I can adapt my designs to certain specifications
		USC 3.2.1	**US 3.1** As a Profit-oriented DHN designer, I want to have an accurate model of a baseline DHN, so that it is easier to design the facility in a more efficient and economic way.
		USC 4.1.1	**US 4.1** As an Efficiency-oriented designer, I want to easily understand how the different parts of the DHN operate, so that I have an accurate view about how the facility is going to work depending on the different configuration that can be proposed.
		USC 4.2.1	**US 4.2** As an Efficiency-oriented designer, I want to have solid demand predictions so that the DHN is designed accurately.
		USC 5.3.1	**US 5.3** As an Informed investor, I want the design of the District Heating Network to be efficient, so that the total amount investment can be reduced.

Table 33: LSP 7 Usage scenarios and user stories considered in MATRYCS

LSP 7 – Usage scenarios and User stories considered in MATRYCS			
UC	#USC	USC	User stories
07	7	USC 1.1.1	**US 1.1** As an Urban Planning Professional, I have innovative ideas and expertise on issues related to deep renovation of urban districts and green infrastructure and I want to better understand the

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

			current situation of the European market and to identify which cities would require my services, so that I can have increased opportunities.
		USC 1.2.1	**US 1.2** As an Urban Planning Professional working with deep renovation of urban districts and green infrastructure, I want to understand which cities are already implementing similar solutions in my area of expertise and what they are implementing, so that I can obtain new knowledges on the current policy targets and these cities' initiatives.
		USC 2.1.1	**US 2.1** As a City department director, I want to understand which targets other urban transport departments across Europe are setting for themselves and which proven good practices exist so that I can use them to define transport related targets and initiatives in my city.
		USC 2.2.1	**US 2.2** As a City department director, I want to quickly and efficiently benchmark my own local transport initiatives against other European cities, so that I can save time in comparison to desk research or knowledge exchange with peers on their own initiative and to achieve better results based on reliable and practical information.
		USC 3.2.1	**US 3.2** As an elected city council member, I want to have aggregated information from across Europe and, so that I can use the data provided not only as a running start for my own policy work but also to identify opportunities to become a European lead city on affordable and sustainable building retrofits.
		USC 4.1.1	**US 4.1** As a European Commission staff, I want to get an up-to-date vision of renovation targets in European cities and their progress against those targets, so that I can understand challenges in increasing the renovation rates in terms of local policy frameworks, needs for technical solutions, finance and investments and local stakeholders' collaboration in order to ensure the call for proposals meets the current needs of European cities.
		USC 4.2.1	**US 4.2** As a European Commission staff, I want to analyse the data to get insights into the pain points of European cities, so that I can ensure the effectiveness of the call for proposals and also to ensure that European initiatives address evident local needs.

Table 34: LSP 8 Usage scenarios and user stories considered in MATRYCS

LSP 8 – Usage scenarios and User stories considered in MATRYCS			
UC	#USC	USC	User stories
UC08_01	2	USC 4.2.1	**US 4.2** As an Worker of Gdynia, I want to get the information about energy and heat usage of public/school building in attractive form so that I can use it it in everyday work in serving the municipality.
		USC 1.1.1	**US 1.1** As a Real estate owner, I want to get all the needed information in one reliable place and free-off-charge so that I can plan an energy renovation for my building and find a subsidy for it.
UC08_02	2	USC 1.1.1 USC 4.1.1	Same two stories as above
UC08_03	3	USC 1.1.1	**US 1.1** As a Real estate owner, I want to get all the needed information in one reliable place and free-off-charge so that I can plan an energy renovation for my building and find a subsidy for it.
		USC 2.1.1	**US 2.1** As a Building administrator and manager, I want to get all the needed information in one reliable place and free-off-charge so that I can plan an energy renovation for my building and find a subsidy for it.

Table 35: LSP 9 Usage scenarios and user stories considered in MATRYCS

LSP 9 – Usage scenarios and User stories considered in MATRYCS

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

UC	#USC	USC	User stories
UC09_01	2	UC 1.1.1	**US 1.1** As an EPC issuer (architect or engineer), I want to check the possible errors in the data entry to EPC softwares so that building owners had trust in calculation of energy consumption and savings related to improvement proposal.
		UC 2.1.1	**US 2.1** As a civil servant of the EPC's registry, I want to be able to develop technical verifications of data and calculations of each EPC registered so that the potential errors are detected and reported to EPC's issuer, therefore increasing the quality of EPC's Regional Administration data sets.
UC09_02	5	UC 1.2.1	**US 1.2** As an EPC issuer (architect or engineer), I want to have a pre-qualification tool and a system of comparing my calculations with nearby certified buildings so that I have an initial estimation of the expected energy efficiency rating, and I can compare it with the calculation obtained.
		UC 4.1.1	**US 4.1** As a potential building buyer or tenant, I want to have any information on the energy efficiency of buildings or dwellings with no energy performance certificate, so that I am able to base my purchase or rental decision on energy issues.
		UC 2.3.1	**US 2.3** As a civil servant of the EPC's registry, I want to offer citizens an informative georeferenced tool for energy performance pre-qualification of all buildings in the region so that building owners become aware of the low energy quality of their constructions, increasing the interest on the implementation of refurbishment works.
		UC 3.1.1	**US 3.1** As a Regional Energy Efficiency planner, I want to use the data of the energy performance certificates of registered buildings so that we can calculate the actual consumption of the building sector in the region, detailing the levels of energy performance by provinces, cities and neighbourhoods.
		UC 3.2.1	**US 3.2** As a Regional Energy Efficiency planner, I want to use the data of the energy performance certificates of registered buildings so that we can prioritize investments of the regional strategy based on the average energy performance of neighbourhoods or cities.
UC09_03	3	UC 2.2.1	**US 2.2** As a civil servant of the EPC's registry, I want to implement an open data web service with all the improvement and refurbishment measures incorporated in the registered energy certificates so that it can inspire other technicians or building owners to undertake refurbishment projects.
		UC 4.2.1	**US 4.2** As a citizen or owner, I want to have an easily accessible and well structure list of potential energy building improvements based on EPC's data so that I can make a better decision about potential energy refurbishments.
		UC 3.3.1	**US 3.3** As a Regional Energy Efficiency planner, I want to use the refurbishment options of the buildings included in the energy performance certificates so that we can develop a better energy efficiency regional strategy, with data-driven investment calculations and deadlines.

Table 36: LSP 10 Usage scenarios and user stories considered in MATRYCS

LSP 10 – Usage scenarios and User stories considered in MATRYCS			
UC	#USC	USC	User stories
UC10_01	3	UC 5.1.1	**US 5.1** As a policy maker working for a government institutions, I want to have a sufficient understanding or supporting information of the energy efficiency situation in the state or local government, so that I can translate these energy savings into economic ones and establish the investment terms and their recovery.
		UC 5.2.1	**US 5.2** As a policy maker working for a government institutions, I want to have a sufficient understanding or lacks supporting information of the energy efficiency situation in the state or local government, so that I can know which building is the priority for investments
		UC 5.3.1	**US 5.3** As a policy maker working for a government institutions, I want to have a sufficient understanding or lacks supporting information of the energy efficiency situation in the state or local government, so that I can know with predictions what will be the energy savings.

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

UC10_02	3	UC 5.1.1 UC 5.2.1 UC 5.3.1	Same three user stories as above.

Table 37: LSP 11 Usage scenarios and user stories considered in MATRYCS

LSP 11 – Usage scenarios and User stories considered in MATRYCS			
UC	#USC	USC	User stories
UC11_01	2	UC 1.1.1	**US 1.1** As a Policy maker, I want to have access to data, to make decisions based on a set of feasible and data-tested options by having access to cross-country analysis of data sets and menus of policy options so that I can make fact- and data-based predictions about outcomes to make informed decisions and convincing arguments.
		UC 2.1.1	**US 2.1** As a policy maker, I want to learn about the effects of certain measures in other countries; to analysis the right conditions for the success of certain policies by having access to a cross-analysis of databases of building stock information and socio-economic profiles of residents and tenancies so that I can do a reality-check about the impact of EU policies.

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

Annex II: User Requirements

ID	Requirement Description	Category	Related Use Cases	Related Services
FR001	MATRYCS should monitor the results to verify energy and cost savings of executed preventive maintenance measures.	Analytics	UC01_01 UC01_02	s.1.4
FR 002	MATRYCS should calculate and monitor sustainability and thermal aspects for facility performance improvement measures.	Analytics	UC01_02 UC04_02	S1.1, S1.2
FR 003	MATRYCS should provide a tool for the identification of possible renovation actions to improve energy efficiency in buildings.	Analytics	UC02_01	S2.2, S4.2
FR 004	MATRYCS should offer a cost assessment tool for the proposed renovation solutions to buildings.	Analytics	UC02_01	S2.2, S4.2
FR 005	MATRYCS should provide a regression model based on historical data.	Analytics	UC03_01 UC06_01	S4.1
FR 006	MATRYCS should provide a statistical engine for method selection.	Analytics	UC03_01 UC06_01	S4.1
FR 008	MATRYCS should provide a forecasting tool for energy related data.	Analytics	UC04_01 UC06_03 UC05_01 UC06_04	S1.1, S1.3, S1.5
FR 009	MATRYCS should provide reports on a facility's behaviour.	Analytics	UC04_01	S1.1, S1.5
FR 010	MATRYCS should provide daily forecast of PV energy production.	Analytics	UC05_01	S1.1
FR 011	MATRYCS should provide energy matching optimization functionality.	Analytics	UC05_03	S1.1 + S1.2

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
FR 012	MATRYCS should perform aggregations on the datasets.	Analytics	UC07_01	S3.1
FR 013	MATRYCS should perform comparisons between available datasets	Analytics	UC07_01 UC09_02	S3.1, S3.2
FR 014	MATRYCS should provide an investment prioritization tool, along with the functionality to simulate scenarios and establish the best-case scenario.	Analytics	UC10_01	S4.2
FR 015	MATRYCS should provide estimations based on the EPBD.	Analytics	UC08_03 UC09_02	S1.3
FR 016	MATRYCS should compare the performance of different buildings.	Analytics	UC08_01 UC08_02 UC09_02	S0.2, s1.3, s3.2
FR 017	MATRYCS should support basic Energy Efficiency Measures calculations.	Analytics	UC02_01 UC03_01 UC03_02 UC06_01 UC06_02	S4.1, S2.2
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
FR 019	MATRYCS should support risk assessment service.	Analytics	UC10_01 UC08_02	S4.2, S1.3
FR 020	MATRYCS should support comparisons between energy efficiency actions.	Analytics	UC10_01	S4.2
FR 021	MATRYCS should support a portfolio analysis engine, based on multiple criteria, in order to support policy making.	Analytics	UC11_01	S3.3

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
FR 022	MATRYCS should provide the selection of building DT services from a list of services on the Digital Twin.	Analytics	UC00_02	S0.1
FR 023	MATRYCS should support statistical analysis of numerical datasets.	Analytics	General	
FR 024	MATRYCS should support data analytics with both private and public datasets.	Analytics	General	
FR 025	MATRYCS should generate EPC based estimations at urban level.	Analytics	UC09_02	S3.2
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 UC04_01 UC04_02 UC05_02 UC05_03 UC06_01 UC07_01 UC10_01 UC11_01	S1.2, S1.4, S4.1, S1.1, S1.5, S3.1, S3.3, S4.2
FR 027	MATRYCS should support the upload of datasets by users.	Collection	UC01_01 UC01_02 UC03_01 UC03_02 UC04_01 UC04_02 UC05_02 UC05_03 UC06_01 UC06_02 UC06_03 UC08_01	S1.2, S1.4, S4.1, S0.2, S1.1, S1.5
FR 028	MATRYCS should support 2D/3D data collection.	Collection	UC00_01	S0.1
FR 029	MATRYCS should support interoperability with APIs for data export.	Collection	General	
FR 030	MATRYCS should provide anonymization 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	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
	from different sources.			
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 UC10_01 UC11_01	S1.1 , S1.2, S3.3, S4.2
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, S0.2
FR 036	MATRYCS should support data import from .xml files.	Collection	UC09_01	S3.2
FR 037	MATRYCS should provide a quering service towards the digital twin.	Compatibility	UC02_01 UC06_04 UC08_02 UC09_02	S2.2, S4.2, S1.3, S3.2, S0.2
FR 038	MATRYCS should support connection to GUIs	Compatibility	General	All services
FR 039	MATRYCS should support connections to web services and APIs.	Compatibility	UC05_01 UC05_02 UC08_01	S1.1, S0.2
FR 040	MATRYCS should combine existing data and user-uploaded data from different sources.	Compatibility	UC07_01 UC08_01 UC09_02	S3.1, S0.2, S3.2
FR 041	MATRYCS should support different types of datasets (historical, live, etc.).	Compatibility	UC00_01 UC00_02	S0.1
FR 042	MATRYCS should utilize the digital twin.	Compatibility	UC00_01 UC00_02	S0.1
FR 043	MATRYCS digital twin functionality should be interoperable with a	Compatibility	UC00_01 UC00_02	S0.1

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
	selection of analytics tools.			
FR 044	MATRYCS should perform data sources integration in order to develop a building's digital twin.	Integration	UC00_01	S0.1
FR 045	MATRYCS should use an advanced ontology.	Integration	UC00_01 UC00_02	S0.1
FR 046	MATRYCS should enable the user to create a new project/activity.	Interaction	UC01_01 UC03_01 UC06_01	S1.4, S4.1
FR 047	MATRYCS should support building identification operation.	Interaction	UC02_01	S2.2, S4.2
FR 048	MATRYCS should support user interaction with the platform, enabling data editing.	Interaction	UC02_01	S2.2, S4.2
FR 049	MATRYCS should support filtering for visualization tools.	Interaction	General	
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 UC04_02	S1.4 + S1.1
FR 053	MATRYCS should provide error messages in case of false user operation.	Notification	General	
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
FR 056	MATRYCS should provide data categorisation tools.	Reporting	UC09_02	S1.3, S3.2

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
FR 057	MATRYCS should be able to report analysis results.	Reporting	UC01_01 UC05_01 UC05_02 UC05_03 UC06_03 UC08_01 UC09_01 UC09_02 UC05_04	S1.4, S1.3, S1.1, S1.2, S1.5, S0.2, S3.2, S2.1
FR 058	MATRYCS should be able to create a list of warnings and possible anomalies.	Reporting	UC01_01 UC04_02 UC05_02	S1.1, S1.4
FR 059	MATRYCS should be able to create a reliable implementation action plan for preventive maintaintance (e.g. in the form of suggestions for improvements).	Reporting	UC01_01	S1.1, S1.4
FR 060	MATRYCS should create the corrective measures to properly mitigate risks.	Reporting	UC01_01	S1.1, S1.4
FR 061	MATRYCS should provide reports on energy savings and cost savings.	Reporting	UC03_02 UC06_02 UC10_01	S4.1, S1.4, S4.2
FR 062	MATRYCS should provide reports on emission savings.	Reporting	UC03_02 UC06_02	S4.1
FR 063	MATRYCS should report the estimated PV performance.	Reporting	UC05_02	S1.1
FR 064	MATRYCS should provide suggestions for increasing self-consumption and energy matching optimization.	Reporting	UC05_03	S1.1, S1.2
FR 065	MATRYCS should provide information about cities' emissions and energy consumption.	Reporting	UC07_01	S3.1
FR 066	MATRYCS should provide specific improvement actions for climate change assessment.	Reporting	UC07_01	S3.1

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
FR 067	MATRYCS should report energy efficiency interventions and actions based on multiobjective analysis.	Reporting	UC11_01	S3.3
FR 070	MATRYCS should provide advanced visualization tools to show the generated results.	Visualization	General	
FR 071	MATRYCS should provide visualizations of indicators or opportunities through maps or charts.	Visualization	UC07_01 UC08_01 UC08_02 UC09_02 UC10_02 UC02_02	s0.2, S3.1, S1.3, S3.2
FR 072	MATRYCS should be able to visualize energy performance information.	Visualization	UC08_01 + UC09_03 + UC02_02	S0.2, S3.2,S1.3
FR 073	MATRYCS should provide classification of the generated solutions.	Visualization	UC10_01	S4.2
FR 074	MATRYCS should provide a graphical representation of optimal energy efficiency portfolios.	Visualization	UC11_01	S3.3
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 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		
NFR 005	MATRYCS should provide time-efficient analytics methods.	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	Related Use Cases	Related Services
NFR 006	MATRYCS should utilize 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 harmonization 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 persistance to avoid data loss	Reliability		

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
NFR 016	MATRYCS should support different authorization access level for different users.	Security		
NFR 018	MATRYCS must ensure the integrity of the customer account information.	Security		
NFR 019	MATRYCS should not return restricted pages to unauthorized 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 visualization 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		

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
NFR 027	MATRYCS should prevent users from inserting invalid input.	Usability		
NFR 028	MATRYCS should provide reports in a standardized 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