Artificial Intelligence for Operation and Maintenance of PV Plants

Deliverable D1.2

Functional specification and system architecture

Lead Beneficiary	ISOTROL
Delivery Date	31/12/2021
Dissemination Level	Public
Status	Released
Version	1.0
Keywords	Specifications, architecture, data, analysis, models

Disclaimer

This work is financed by the ERDF - European Regional Development Fund through -the Operational Programme for Competitiveness and Internationalisation COMPETE 2020 under the Portugal 2020 Partnership Agreement within project AI4PV, with reference POCI-01-0247-FEDER-111936 – and Spain's CDTI Multi-regional Operational Programme 2014-2020. International collaborative project EUR 2020058 with the seal of the AI EUREKA CLUSTER.

Document Information

Project Acronym	AI4PV
Work Package	WP 1
Related Task(s)	T1.2
Deliverable	D1.2
Title	Functional specification and system architecture
Author(s)	Rubén González (Isotrol), Daniel Jiménez (Isotrol), Christian Verrecchia (EDP NEW), Louelson Costa (INESC TEC)

Revision History

Revision	Date	Description	Reviewer
0.1	04/10/2021	Outline of report content	Isotrol
0.2	30/11/2021	Full draft of all content	Isotrol
0.3	10/12/2021	First Version	EDP, INESC TEC
0.4	17/12/2021	Second Version	Isotrol
1.0	30/12/2021	Final Version	EDP

EXECUTIVE SUMMARY

This deliverable contains a description of the AI4PV system expanding the initial definition provided as a result of the definition of Use Cases (UCs) previously defined in Deliverable D1.1. This definition focuses on the aspects that were not addressed as a part of the work done in T1.1 and D1.1, these are: i) the description of what information from a potential environment would be necessary for the proper development and deployment of the AI4PV solution; ii) the design of the technical and functional architecture for the deployment of the project, and iii) the definition of the requirements of the system as a whole.

The contents of this documents have been obtained in a process of analysis of mechanisms for the deployment of analytical systems such as the proposed one for AI4PV, the analysis of the information available from PV plants which could be used as pilots for the validation of the project and the high-level description of potential use environments. From this point, an in-depth analysis of cloud solutions and implementations systems has been performed, process in which Google Cloud Platform has been selected as the tool for the deployment of the project; and, after that, the proper technological architecture has been defined for the project using services compatible with said infrastructure.

Finally, a list of specifications to characterise and clarify the scope of the project is presented, a list which has been created using the UCs and the previous knowledge from the companies in the consortium to finalize in a list of what to be expected as a result of the project.

TABLE OF CONTENTS

Executive summary ... 3
Table of contents ... 4
List of figures ... 5
List of tables ... 6
Abbreviations and acronyms ... 7
Glossary of key terms ... 8
1. Introduction ... 9
1.1 Scope of report ... 9
1.2 Outline of report ... 9
2. Plant description ... 10
2.1 Plant information ... 10
2.2 PLant characterization ... 11
3. System architecture ... 13
3.1 FUnctional architecture ... 13
3.2 Technical architecture and interfaces ... 14
3.3 Cloud integration ... 15
3.3.1 GCP (Google Cloud Platform) ... 15
3.3.2 AWS (Amazon Web Services) ... 15
3.3.3 Cloud provider selection ... 16
3.3.4 Solution's provider security ... 17
4. System specifications ... 18
4.1 Functional requirements ... 18
4.2 Data requirements ... 22
5. Conclusions ... 27

LIST OF FIGURES

Figure 2-1: Temperature coefficients 11
Figure 3-1: AI4PV FUNCTIONAL ARCHITECTURE 13
Figure 3-2: AI4PV Technological ARCHITECTURE 14

LIST OF TABLES

Table 2-1: Plant information 10
Table 2-2: Plant characterization 11
Table 4-1: FR1 User access 18
Table 4-2: FR2 Data cleansing and quality analysis 18
Table 4-3: FR3 Modular and expandable system 19
Table 4-4: FR4 Multi-topology integration 19
Table 4-5: FR5 failure and underperformance detection 20
Table 4-6: FR6 Parameter optimization 21
Table 4-7: FR7 O&M recommendation generation 21
Table 4-8: Plant characterization 22
Table 4-9: Power station data 23
Table 4-10: Inverter data 24
Table 4-11: StringBox data 25

ABBREVIATIONS AND ACRONYMS

Acronym	Meaning
AI	Artificial Intelligence
API	Application Programming Interface
AWS	Amazon Web Service
CMEK	Customer-Managed Encryption Keys
CPU	Central Processing Unit
DC	Direct Current
DLP	Data Loss Prevention
DNS	Domain Name System
DQR	Data Quality Routines
DT	Digital Twin
EBS	Elastic Block Store
EC	Elastic Compute
FR	Functional Requirements
GB	Giga-Byte
GCP	Google Cloud Platform
IoT	Internet of Things
KPI	Key Performance Indicator
ML	Machine Learning
PPC	Power Plant Controller
PV	Photovoltaic
RDS	Relational Database Service
RoI	Return on Investment
SCADA	Supervisory Control and Data Acquisition
SQS	Simple Queue Service
UC	Use Case
VM	Virtual Machine
VPC	Virtual Private Cloud
VPN	Virtual Private Network

GLOSSARY OF KEY TERMS

Machine Learning	Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.
Artificial Intelligence	Artificial intelligence is a wide-ranging branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence.
Digital Twin	A digital twin is a virtual representation that serves as the real-time digital counterpart of a physical object or process.
Fault	A fault is an unpermitted deviation of at least one characteristic property (feature) of the system from the acceptable, usual standard condition.
Failure	Permanent interruption of a system's ability to perform a required function under specified operating conditions.
Fault detection	Determination of faults present in a system and time of detection.
Cloud computing	Cloud Computing is Internet-based computing, whereby shared resources, software, and information are provided to computers and other devices on demand.

1. INTRODUCTION

This document is entitled "Functional specification and system architecture" and it is developed as part of the AI4PV project, and its ultimate goal is to provide a full description of the AI4PV architecture to be deployed in order to collect and analyse operational data from the PV park at the scope of detecting faults and failures and providing recommendation to the O&M team to solve the identified issues.

This document is framed by the Work Package 1 (Project Management) to act as deliverable D1.2, and it will be used to define the overall architecture and ensure that the AI4PV solutions developed in WP2 and WP3 will meet the requirements and set of KPIs defined for the different UCs in Deliverable D1.1. Similarly, this document and the architecture herein presented will ensure compatibility between the system already in place in the PV park under issue and the one envisioned for the AI4PV project.

1.1 SCOPE OF REPORT

This document focuses on the description of the Global specification for data analytics and modelling of PV Plant critical components. Different aspects have been considered in the process, such as specification for data gathering for analysis and model's comparison from SCADA and sensors. In addition to that, and to ensure that said description is achievable, the work here gathered also includes the global technical architecture definition for the AI4PV system.

It contains the main results obtained in the execution of task **T1.2 Functional specification and system architecture**, part of the **WP1 Design of a global architecture & requirements identification**. It includes the main results obtained in the high-level technical design and definition of the project, including the architecture of the system, the interfaces between modules themselves and data sources, and the specific requirements for the complete system.

1.2 OUTLINE OF REPORT

The structure of this report is the following:

- ***Chapter 1*** introduces the scope of the project and the deliverable in its context.
- ***Chapter 2*** provides a brief summary of the analysis of the characterization and descriptive information from PV plants that has been identified as necessary for the proper deployment of the project, and the specific information considered for validation in this stage of the project.
- ***Chapter 3*** focuses on the definition of the architecture designed for the project from two perspectives, functional and technological attending to cloud deployment requirements.
- ***Chapter 4*** includes a list of specifications and requirements identified for the project, from functional information to constraints regarding data provision.
- ***Chapter 5*** includes a summary of the conclusions achieved after finalising the deliverable and T1.2.

2. PLANT DESCRIPTION

This section includes a list of the information which has been identified as necessary in order to ensure a successful development of the solutions, since unclear information or with not enough details can lead to misinformation and misinterpretation of the configuration of the environment, resulting in wrong algorithms detections and not-optimal systems.

2.1 PLANT INFORMATION

In order to characterize and define the scope, specifications, and, in the end, to help define the AI4PV proposed solutions, it is important to define the environment in which the developments will be implemented. Table 2-1 presents an initial and brief description of the plant that is being considered for the deployment and validation of the solutions.

TABLE 2-1: PLANT INFORMATION

Plant name	Monte das Flores
Country of location of the plant	Portugal
Type of inverters - These can be string-inverters, central-inverters or inverters with two or more inverter-modules.	4 Central Inverters. Model EFASOLAR 630
Inverters subcategory - To see if they are multi-MPPT inverters or inverters with Master/Slave configuration.	1 MPPT for each inverter
Name of the inverters in the SCADA - Normally the source data comes with a different name which, through mapping, is renamed to the ones actually used.	INV1.1, INV1.2, INV2.1, INV2.2
Nominal DC power in MW installed in the plant according to the modules - It will serve to see the possible oversizing of the plant.	2.9
Maximum ActivePower in kVA of the inverters according to the manufacturer	630 (each at 50ºC)
Maximum DC voltage in V	830
Nominal cell operating temperature	Range -40ºC to 90ºC
Tracker - Type of connection of the panels, if they are fixed, if they are connected by tracker and in this case if it is one or two axes	Main plant - Fixed

	Single-axis tracking systems are available at the site
Inclination at which the panels are connected	20º facing South
Bifaciality - Type of panels, whether they are bifacial or not.	Main plant – Monofacial (model UP-M310P)
Bifaciality coefficient - In case of bifacial panels.	N/A
Azimuth angle in which the panels are connected	N/A
Temperature correction coefficient - To correct the irradiance received by temperature.	Temperature Coefficients NOCT (°C): 45 ± 2 Temperature Coefficients of Isc (% / °C): 0.05 ± 0.01 Temperature Coefficients of Voc (% / °C): -0.30 ± 0.02 Temperature Coefficients of Im (% / °C): -0.02 ± 0.02 Temperature Coefficients of Vm (% / °C): -0.42 ± 0.03 Temperature Coefficients of Pm (% / °C): -0.43 ± 0.05 **FIGURE 2-1: TEMPERATURE COEFFICIENTS[1]**
Area in m² of the panels - It can be the data per panel.	1.950 m² all surface. 1.752 m² active surface (72 cell, 156 mm x 156 mm)

This information would be mandatory for the development and deployment of any similar solution, since it will allow to characterize information such as the expected power output, the weather effect on power production according to the span and geometry of the PV plant, and how many inverters/strings have to be considered.

As it can be appreciated in the table, the plant proposed for the developments and validation of the project's solutions, is a utility-scale PV park with characteristics which can be found in many other PV plants. These similarities with any other utility-scale PV park ensures that the developments obtained in the context of the project can easily be replicated to others PV plants.

2.2 PLANT CHARACTERIZATION

Table 2-2 exposes information that characterizes the PV plant under issue with a higher level of detail.

TABLE 2-2: PLANT CHARACTERIZATION

Relation between stringbox/combinerbox-strings-modules	Each String corresponds to 20 PV panels. Combiner boxes with 13 Strings, 14 Strings and 16 Strings.

[1] *Datasheet Upsolar, model UP310MP*

	Each inverter has 8 Combiner boxes. In total each inverter has 117 Strings with 2340 PV panels.
During the life cycle of the plant, has it undergone any modification (if modules have been added, changes of manufacturer).	No
Elements that provide meteorological signals (pyranometers, weather station or satellite).	Pyranometer Horizontal Pyranometer 20º. Wind speed and direction Ambient Temperature. Atmosphere Pressure. Pluviometer. Pireliometer.

The information here presented is also required for a proper description of the plant. It allows to identify how the data reflects the information over the real PV plant and how to link the data with the energy flow in the complete process from PV modules to the injection into the grid.

As it can be seen, the information in Table 2-2 mainly reflects aggregations of elements in different layers of the PV plant and how the weather information for these levels of aggregation are available.

Other information which could be considered depending on the specific environment and the characteristics of the plant are the following:

- Association Tracker - Strings or Tracker - String Box.
- If there are several meters, which one is considered the **reference**.
- PPC signals.
- Alarm data (not aggregated)
- Manufacturer specifications regarding the efficiency curve of the inverter.
- Which pyranometer is linked to each inverter.

3. SYSTEM ARCHITECTURE

3.1 FUNCTIONAL ARCHITECTURE

AI4PV proposes the design and development of a system capable of combining the best advantages of Machine Learning (ML), Artificial Intelligence (AI) and system and components modelling in order to successfully detect failures and underperformances in PV systems. In order to achieve such an ambitious approach to the analysis of the operation of PV plants using only automated data analytics, a complex system structure has to be defined, in which several processes take place in an orchestrated way. In this stage of the project, and according to the definition of UCs addressed in the project and defined in Deliverable D1.1, an initial dataflow and process tree has been defined, which is represented in Figure 3-1

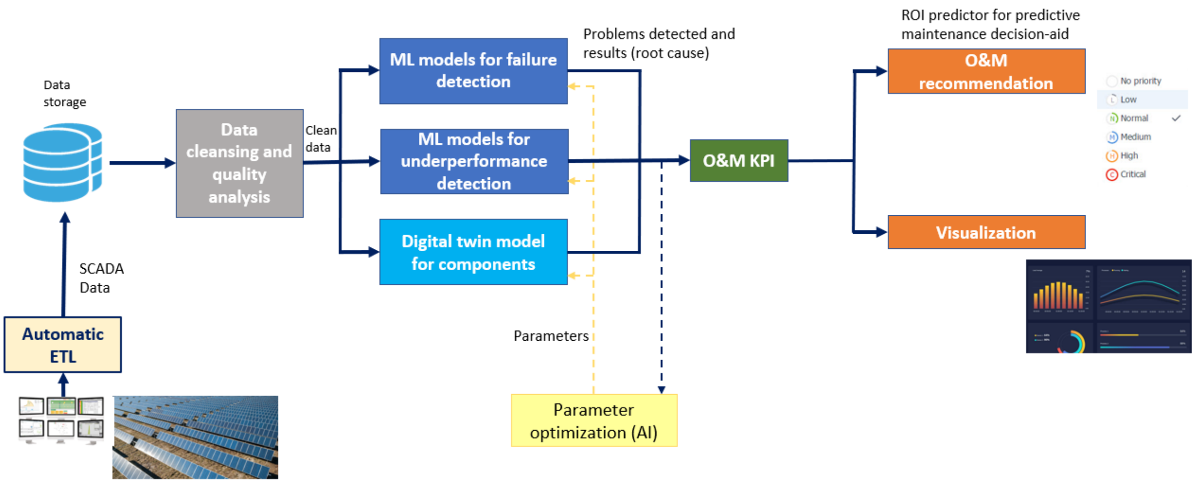

FIGURE 3-1: AI4PV FUNCTIONAL ARCHITECTURE

As it is shown, in order for the system to be capable of working in real environments, an automatic tool for the extraction of SCADA data must be developed and prepared to work as a first interface between the AI4PV system and the data sources for the execution of analysis.

Data extracted from the field, would later pass through a stage of preparation, in which their quality will be analysed, and problems related to incomplete, wrong or null data will be addressed, in order to make sure that only relevant and valuable data is used by the algorithms in charge of the detection of problems, failures and underperformances.

Once problems have been identified, analysed and cleaned, only valuable data and Key Performance Indicators (KPIs) will be used to categorize the criticality and importance of each of the problems, to finally feed a recommendation engine that prioritizes failures and problems according to these KPIs, providing end users a list of actions to be conducted in the best possible order to maximize power production in the long term and minimize unavailability.

Finally, the parameters for failure and underperformance detection will be constantly updated and the algorithms optimized using the results obtained such as success in the detection and accuracy.

3.2 TECHNICAL ARCHITECTURE AND INTERFACES

For the implementation of the functional description of the system described in the previous section, a cloud powered architecture has been defined, as illustrated in Figure 3-2.

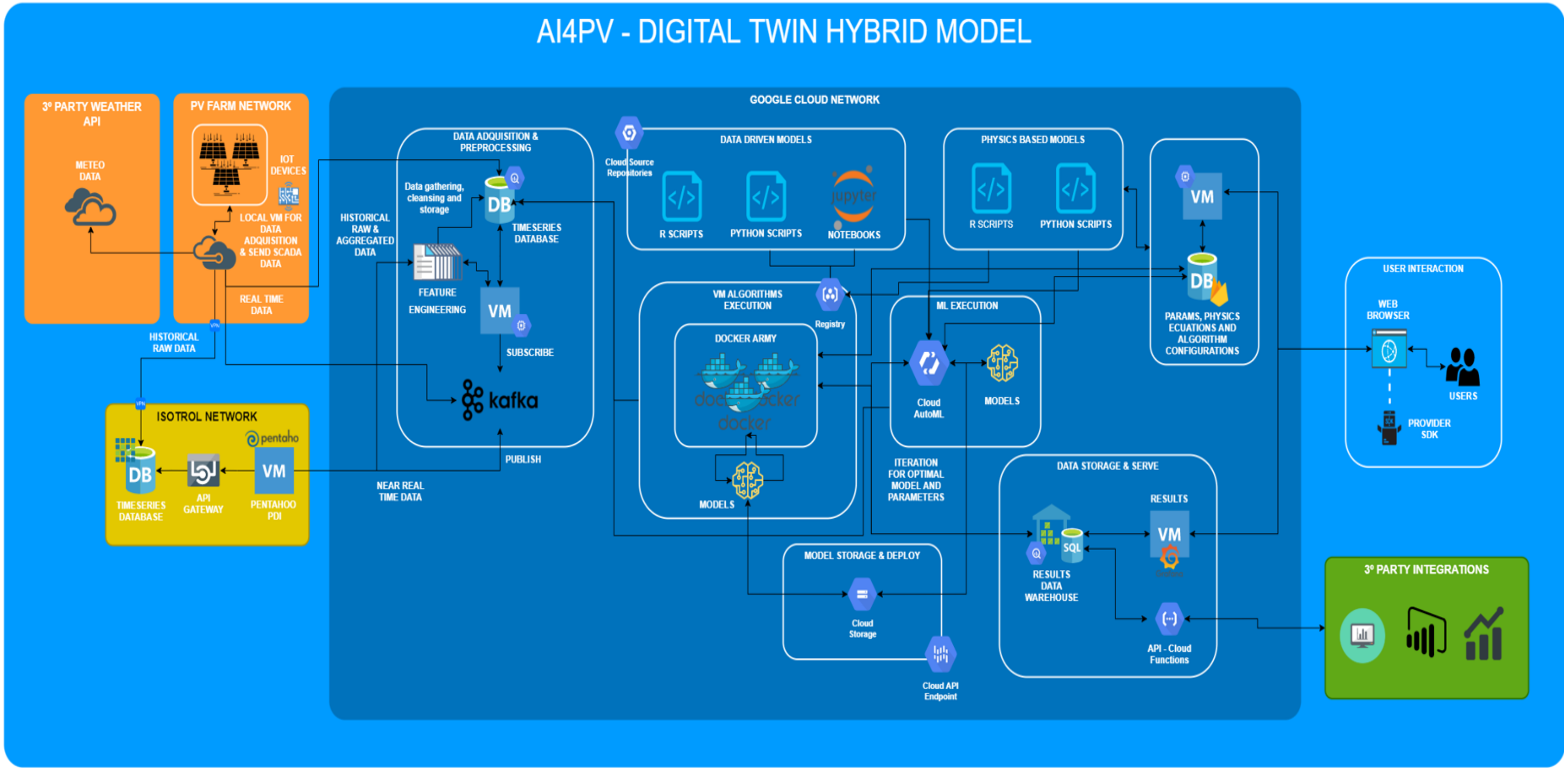

FIGURE 3-2: AI4PV TECHNOLOGICAL ARCHITECTURE

COMPETE 2020 PORTUGAL 2020 @CDTIoficial European Union European Regional Development Fund

The system will be implemented using a cloud infrastructure, capable of perfectly connecting with external data providers, such as the SCADA system (from which the PV operational data will be gathered) and other third-party data providers, as meteorological data. The resulting tool will be capable of integrating visualization layers for the view of the results of the different analysis, using graphs and other elements.

It has been proposed that a unified storage is used for the extraction of data directly from the PV plant, either by the use of VPNs or gateways to send the data from the source to the storage, which will be hosted by Isotrol. From there, the data will be shared with the consortium. Each of the partners will develop a set of tools for the implementation of analytics, which will be implemented in Virtual Machines (VMs). Each VM can be hosted in different environments: it can either run locally on a partner's device for confidentiality issues or it can be implemented in the AI4PV architecture itself but, in the end, the architecture portrays a view of an implementation that will be later conceptualized according to the real environment and specifications of the environment.

3.3 CLOUD INTEGRATION

Once the solution has been proposed, different cloud services providers have been analysed in order to implement it. These providers base their services on the web, taking advantage of all its potential for the deployment of cloud architectures, which implies that no local infrastructure needs to be maintained, being everything in the providers' networks. For the specific situation of the project, two different providers have been analysed: GCP (Google Cloud Platform) and AWS (Amazon Web Service).

3.3.1 GCP (GOOGLE CLOUD PLATFORM)

As a result of an analysis carried out to identify different features of this provider, the services that should be used to implement AI4PV architecture are the following:

- Computation: Google Compute Engine and Google App Engine
- Storage: Google Cloud Storage
- Networking: Google Cloud DNS
- Databases: Google Cloud Datastore and Google Cloud Bigtable
- Data warehouse: Google Bigquery
- Serverless: Google Cloud Functions, Cloud Scheduler, Cloud EndPoints
- Container images: Cloud Build, Container Registry
- Results visualization: Data Studio
- Messages: Cloud PubSub

3.3.2 AWS (AMAZON WEB SERVICES)

Once this provider has been analysed, the services that should be used to implement this solution are the following:

- Computation: Amazon Elastic Compute Cloud (EC2)
- Storage: Amazon Simple Storage Service (S3) y Amazon Elastic Block Store (EBS)
- Networking: Amazon Virtual Private Cloud (VPC)
- Databases: Amazon Relational Database Service (RDS) y Amazon DynamoDB
- Data warehouse: Google Bigquery
- Serverless: AWS Lambda, AWS Job Scheduler, API Gateway
- Container images: AWS CodeBuild, EC2 Container Registry
- Results visualization: QuickSight
- Messages: AWS SQS

3.3.3 CLOUD PROVIDER SELECTION

For the development and deployment of the project, GCP has been selected as the main service provider, since it already offers a series of advantages that can be used to get all the performance expected for AI4PV project. AWS is also a good provider, but it has been decided that, for this solution, GCP can provide to us better solutions. The main features from GCP that have justified this choice are reported below.

Custom machines

The Google Cloud Platform allows to configure the optimal combination of virtual memory and CPU for any workload, so the is no longer need to over-provision to fit a provider's rate plan. This saves more and improves the price/performance ratio by up to 50%, depending on the workload.

Better pricing and data restoration

Google Cloud Storage Nearline offers data availability in less than a second and provides high performance for fast data restoration. In fact, it's so fast that many companies use it simply as their only level of storage. GCP is the only provider with clear rates: $0.01 per GB/month for storage, $0.01 per GB for restore.

User Load Management

Its integrated load balancer is part of a globally distributed system to connect customers to the infrastructure. It is the same system used in Google products such as Maps, Gmail and search. Its automatic scaling features are designed to tolerate extreme traffic surges and smoothly scale in seconds from no traffic at all to millions of requests per second. This improves application performance - no matter how many users appear at once.

Increased Speed

Compute Engine instances take between 40 and 50 seconds to start: about a fifth of what other clouds take. This means that the service capacity of the application can be extended very quickly in response to incoming traffic.

3.3.4 SOLUTION'S PROVIDER SECURITY

At Google Cloud, they've set a high bar for what it means to host, serve, and protect customer data. Security and data protection are fundamental to how they design and build their products. They start from the fundamental premise that GCP customers own their data and control how it is used. The data a customer stores and manages on GCP systems is only used to provide that customer with GCP services and to make GCP services work better for them, and for no other purpose.

They have robust internal controls and auditing to protect against insider access to customer data. This includes providing their customers with near real-time logs of Google administrator access on GCP; GCP is the only major cloud to provide this level of access transparency. In addition to continuous security monitoring, all customer data stored in GCP is encrypted at rest and in transit by default. Customers can also choose to manage their own encryption keys using GCP's Cloud Key Management Service, a feature commonly referred to as "customer-managed encryption keys (CMEK)."

They also enable their customers to monitor their own account activity. They provide reports and logs that make it easy for a customer's administrator to examine potential security risks, track access, analyse administrator activity, and much more. Administrators in our organization can also leverage Cloud Data Loss Prevention (DLP) capabilities to protect sensitive information. DLP adds a layer of protection to identify, redact, and prevent sensitive or private information from leaking outside of an organization. Additionally, our administrators can also enforce policies over mobile devices in their organization, encrypt data on devices, and perform actions like remotely wiping or locking lost or stolen devices.

Furthermore, they undergo independent, third-party audits and certifications to verify that our data protection practices match our commitments.

Google's global infrastructure coverage is the best encrypted and backed up in different parts of the world and connectivity between each of them, making it the best option for having the service in the cloud.

It implements in an infrastructure protected by more than 500 of the best experts in information, application and network security. Google's security model is a process developed over 15 years of experience to keep customers protected when they use Google applications such as Gmail and G Suite. With the Cloud Platform, applications and data benefit from the same comprehensive security model.

As a summary, the use of Google Cloud will provide a robust system for the deployment of the solution proposed by the project that will ensure the safety of the data hosted while providing flexible and accessible services for the developers in the consortium.

4. SYSTEM SPECIFICATIONS

In this section of the document, a description of the requirements that have been defined for the project is presented. Two categories of requirements are presented, these are, **functional requirements (FR)**, meaning what the system will do, and **data** requirements, focused on what data would be necessary in order to properly implement said functional requirements.

4.1 FUNCTIONAL REQUIREMENTS

The following tables include a description of the functional requirements: these requirements describe what the AI4PV system will do once it is developed, including a high-level description of its features and what end users can expect of it:

TABLE 4-1: FR1 USER ACCESS

ID	FR1		
Name	User access		
Priority	☒ High	☐ Medium	☐ Low
Requirement description			
Users will have access to the system through a **web access screen**, where they must enter their username and password to log in. Once they have entered the system, the user will have at his disposal the different tools or modules configured for his profile from the user management, as well as the possibility of interacting with the system or logging out. The system will only allow registered users to access the platform and will include password modification services, which could be linked to specific predefined accounts, such as Google accounts.			

TABLE 4-2: FR2 DATA CLEANSING AND QUALITY ANALYSIS

ID	FR2		
Name	Data cleansing and quality analysis		
Priority	☒ High	☐ Medium	☐ Low
Requirement description			
The platform will be able to process the data acquired from SCADA or Internet of Things (IoT) sensors in PV power plants and do a proper processing of the information, in order to ensure that			

the data used for the identification of issues and the generation of recommendations works with real, meaningful and relevant data. Data cleansing will be implemented using a modelling approach based on solving equations which define the behaviour of each asset and determine if the data is consistent or not, according to its modelled behaviour. The models will be based on electrical and empirical equations that follow the specifications and characteristics of each asset to be analysed. More detailed mathematical models, Sandia labs models, linear regression, or systems as Simulink will also be considered. Finally, the integrity of the data will be assessed using Data Quality Routines (DQRs), aim at providing good quality timeseries data before conducting any additional performance analysis (such as failure detection and classification).

TABLE 4-3: FR3 MODULAR AND EXPANDABLE SYSTEM

ID	FR3		
Name	Modular and expandable system		
Priority	☐ High	☒ Medium	☐ Low
Requirement description			
The AI4PV platform will be designed in order to allow the simple integration of new developments and modules for the detection of specific failures or problems. This integration will allow that data inputs are directly accepted by new modules deployed in the platform and that the results of the processing can easily be transformed into recommendations for O&M. The approach will begin in the design phase, in which data structures will be described as general as possible, making it easy for new developments to be adapted into the specific data structure of AI4PV. Then, the outputs of the analytic tools (and inputs to the recommendation engine) will have a common, easy to understand, format, allowing the simple transformation into recommendations and ensuring the modularity and expandability of the system.			

TABLE 4-4: FR4 MULTI-TOPOLOGY INTEGRATION

ID	FR4		
Name	Multi-topology integration		
Priority	☐ High	☒ Medium	☐ Low
Requirement description			
Developments must be versatile, and integrable with plants with different topologies. Instead of providing a solution specifically designed to be implemented in a given context, the AI4PV project will focus on the development of a flexible and accessible platform, which allows to add different			

modules and rules to the optimization system according to the topology and distribution of assets in different PV plants, by a simple parametrization of the developments. Changes in configurations of the PV modules, inverter topologies, technology of PV modules, and the connection in combiner boxes, among others, will be considered. The system will be structured following a **cloud** architecture and **service provision** concept, which will ease the implementation of a solution flexible with integrable with different environments.

TABLE 4-5: FR5 FAILURE AND UNDERPERFORMANCE DETECTION

ID	FR5		
Name	Failure and underperformance detection		
Priority	☒ High	☐ Medium	☐ Low
Requirement description			

The AI4PV platform will provide a set of services to users focused on the improvement of operation and maintenance in PV plants. This improvement will be based on the identification and early detection of problems and failures in PV operational data, using ML and digital twin (DT) technologies to identify deviations from normal operation.

The detection of failures and underperformances which could indicate a problem will focus the scope on a set of specific failures, since the list of potential failures in PV plants can be overwhelming to address, and, in most situations impossible to detect due to the lack of data and sensors.

The specific list of failures modes which have been identified so far is the following:

Solar Field:

- String & stringbox disconnection
- Tracker blocking
- Tracker misalignment
- Panel ageing (time degradation)
- Sensor (pyranometer, current, power, etc.) malfunction

Inverter:

- Inverter shutdown
- Temperature disconnection
- Maintenance stop
- Late awakening
- Clipping
- MPPT non-optimal

• Out of normality behaviour **Transformer:** • Open circuit • Short circuit • Out of normality behaviour This list is preliminary and might be slightly modified if any of the problems is not detectable due to the lack of previous events that allow the training of models, or the identification of new relevant failures that should be detected.

TABLE 4-6: FR6 PARAMETER OPTIMIZATION

ID	FR6		
Name	Parameter optimization		
Priority	☒ High	☐ Medium	☐ Low
Requirement description			
The AI4PV system will include a module for the automatic optimization of parameters in the execution of the different analytic tools included in its context. Parameters include information that defines the context in which the algorithms are executed, affecting the way in which the data is gathered, analysed and, in the end, the results of the analysis is produced. Sometimes, the identification of the parameters is clear and leaves no room for error, however, in some situations, it is hard to identify correct values. Parameters provided by PV modules or inverter manufacturers, e.g., may change due to time-related degradation of the systems, or repairs and piece replacements, might change values and, in the end, analysis results might be less accurate than expected. In order to increase the quality of the problem detection and recommendation generation, a module to analyse the results of the analysis is provided (to compare the real and predicted values, and any other relevant inputs), in order to update parameters and improve the accuracy of the detections and predictions of the AI4PV platform. This will affect parameters related to the plant configuration and for the characterization of the assets and will be applied to both ML and DT processing.			

TABLE 4-7: FR7 O&M RECOMMENDATION GENERATION

ID	FR7
Name	O&M RECOMMENDATION GENERATION

Priority	☒ High	☐ Medium	☐ Low
Requirement description			

The platform will provide a prioritized list of O&M recommendations as a result of the post-processing of the detection of failures and underperformance. This list will help to schedule O&M activities, maximising the RoI (Return on Investment) by optimizing both the availability of the plant, and the costs related to maintenance, repairment or warranty claiming.

The system must provide a brief explanation of the reasons which infer each of the recommendations, and use information such as the potential cost or impact in production caused by the problem or by ignoring a given recommendation in order to define the criticality of the problem.

To sum it up, the AI4PV platform will determine the best action to take when a problem or degradation has been detected, depending on the number of elements and the power or energy affected, proposing the best alternative between early predictive action or waiting for corrective action.

The recommendations will be presented using text to explain the actions to perform and some sort of representation of the "impact" or "criticality" of the situation, using KPIs or past information in order to define the specific values.

4.2 DATA REQUIREMENTS

This section includes a list of the different data that has been defined as necessary for the proper detection of problems and failures in the context of the project. As it has been done in Section 2 of this document, the table has been filled with information available from the plant which is being considered for the deployment of the developments of the project.

First of all, the plant characterization is key, since it will allow to identify the normal operation conditions of the PV plant under study, which is shown in Table 4-8:

TABLE 4-8: PLANT CHARACTERIZATION

Device	Subdevice	Parameter	Unit	YES/NO
Manufacturing characteristics	Transformer			YES
	Inverter datasheet			YES
	Pyranometer			YES
	Anemometer			YES
	Location	Evora		

Technical details of the plant	Latitude	38,542		
	Longitude	-7,961		
	Altitude	227	mt	
	Power Capacity	2.9	MWp	
Distribution	Plant Hierarchy			YES
	Diagram line connection			YES
Other consideration	Data acquisition CSV format			YES
	Maintenance report			YES
	Known issues			YES

With this initial description, parameters will be defined for the specific definition of the detection mechanisms. In addition to those initial parameters, the list of relevant signals for the execution of analysis and the evaluation of the performance is a key element to consider in this stage of the project. In Table 4-9,Table 4-10 and Table 4-11, a clear definition of the most important data which has been identified so far, is presented.

TABLE 4-9: POWER STATION DATA

Device	Subdevice	Parameter	Unit	YES/NO
Power station/Meter	General	Status		N/A
		Active Energy	kWh-MWh	YES
		Active Power	kW	YES
		Reactive Power	kVAr	YES
		Daily Active Energy	kWh	YES
		Apparent Power	kVA	YES
		AC Voltage	kV	YES
		AC Current	A	YES
		Grid Frequency	Hz	YES
	Cumulative	Cumulative Irradiance	kWh	YES
		Daily Cumulative Active Energy	kWh	YES

		Daily Cumulative Reactive Energy		NO
	Consumed	Consumed Active Power		NO
		Consumed Reactive Power		NO
	Delivery	Delivery_line_to_line_voltage_U_{RS}	kV	YES
		Delivery_line_to_line_voltage_U_{ST}	kV	YES
		Delivery_line_to_line_voltage_U_{TR}	kV	YES

TABLE 4-10: INVERTER DATA

Device	Subdevice	Parameter	Unit	YES/NO
Inverter	General	Status		YES
		Active Energy	kWh	YES
		Active Power	kW	YES
		Reactive Power	kVAr	YES
		Daily Active Energy	kWh	YES
		Power Factor		YES
		Frequency	Hz	YES
		Apparent Power	kVA	YES
		Tanphi		YES
		cosphi/cosphi curve		YES
	AC Voltage	Phase R-S	V	YES
		Phase S-T	V	YES
		Phase T-R	V	YES
	AC Current	Phase R	A	YES
		Phase S	A	YES
		Phase T	A	YES

	DC	DC Voltage	V	YES
		DC Current	A	YES
		DC Power	kW	YES
	Power electronics	Status		YES
		Power Electronic Temperature	ºC	YES
		Fault		N/A
		Warning		N/A

TABLE 4-11: STRINGBOX DATA

Device	Subdevice	Parameter	Unit	YES/NO
Stringbox	General	Status		N/A
		DC Voltage	V	YES
		DC Current string	A	YES
		DC Power string	kW	YES
		Fuse String		N/A
		Temperature	ºC	YES

As it can be appreciated, in the specific case of the plant which is being considered for the deployment of the solution in the context of the project, most of the information and data is available, which is a guarantee that the detection capabilities of the system will be able of identifying plenty (if not all) of the failure mode which have been selected in the context of the project.

In addition to the list of relevant signals, two other requirements have been identified to ensure the proper performance of the tool:

- **1 year of data (minimum), ideally 2 years:** In order to identify normal operation conditions, and to allow the proper training of machine learning models, a minimum of 1 year of data is required. This is to ensure a proper characterization of seasonal effect in different signals and behaviours. 2 years is better since it allows to identify anomalies and better define normal operation conditions. In the framework of AI4PV data from 2018 to early 2022 are available.
- **5-minutes granularity:** In order to detect problems in PV systems, a minimum time frame of 5 minutes per measurement is required, since larger time-span granularities may be the reason for problems and deviations to go unnoticed. Usually, a granularity below 5-minuntes

do not provide a better detection of problems but heavily affects the storage and computing necessities, reason for what 5 minutes is considered the best granularity for the system.

5. CONCLUSIONS

This deliverable includes a clear description of the scope of the project, which, in addition to the definition of the UCs (outlined in Deliverable D1.1), complete the definition stage of the project AI4PV. The specific contents of this document are the results of the process of architecture and specification definition.

As a result, a system, based on a cloud powered architecture, has been proposed, in which raw data will be used as an input, which will later be cleansed and prepared for the execution of advanced algorithms for the early detection of failures in PV assets, using the combined power of ML and DT. Later, those detections will allow end-users to optimize O&M activities using a set of recommendations, provided by the tool according to the failures detected and the impact that each problem could have on the operation of the PV plant under study. Google Cloud will be the tool used for the deployment of the solution.

A set of specific functional specifications have been presented, which includes the complete scope of the project, these are:

1. FR1 USER ACCESS
2. FR2 DATA CLEANSING AND QUALITY ANALYSIS
3. FR3 MODULAR AND EXPANDABLE SYSTEM
4. FR4 MULTI-TOPOLOGY INTEGRATION
5. FR5 FAILURE AND UNDERPERFORMANCE DETECTION
6. FR6 PARAMETER OPTIMIZATION
7. FR7 O&M RECOMMENDATION GENERATION

Finally, data requirements for the execution of the algorithms and the implementation of the functional requirements are presented. The list presented includes a list of all the required information for a clear detection of all the problems whose detection is expected. The plant selected for validation includes most of the information, which ensures that, if not all of it, most of the scope of the project will be easy to implement.

COMPETE 2020 PORTUGAL 2020 CDTI @CDTIoficial European Union European Regional Development Fund