D9.1 - Platform architecture designed and released

September 2024

SUITE5

Table 1: Document information

Document Information	
Project name:	iAMP-Hydro
Project title:	Intelligent asset Management Platform for Hydropower operation & maintenance
Project number:	101122167
Start date:	October 2023
Duration:	36 months

Table 2: Dissemination level of this report

Dissemination level of this report	
PU	Public **X**
PP	Restricted to other programme participants (including the Commission Services)
RE	Restricted to a group specified by the consortium (including the Commission Services)
CO	Confidential, only for members of the consortium (including the Commission Services)

Document History

Version	Date	Document history/approvals
0.10	26/04/2024	Draft ToC
0.20	30/04/2024	Draft Chapter 1
0.30	14/05/2024	Draft Chapter 2, 3, 4
0.40	17/06/2024	Updated chapter 3,4 with partners contributions
0.50	19/07/2024	Updated and sent for partners contributions
0.60	30/07/2024	Updated with partners contributions
0.70	02/09/2024	Updated with partners contributions, draft chapter 5
0.80	17/09/2024	Updated as per internal review and final draft sent for peer review
0.90	30/09/2024	Final document updated as per peer review comments
1.00	30/09/2024	Provided to the PC for submission

ACKNOWLEDGMENT & DISCLAIMER

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101122167.

EXECUTIVE SUMMARY

The deliverable at hand, titled: D9.1 - Platform architecture designed and released, outlines the comprehensive process followed and resulting outcomes of T9.1 and T9.2, activities which are pivotal to the development of the iAMP-Hydro platform. In the context of T9.1, this deliverable focuses on defining the project's use cases and eliciting the user requirements through all consortium partners' participation. To address the objectives of T9.2 this document translates these extracted requirements into a detailed architecture design for the iAM Platform; which is defined in detail in this deliverable, along with the description of the platforms' intra-components and their interfaces. By engaging end-users at the projects initial phases and meticulously designing the system architecture, it is ensured that iAMP-Hydro will meet the diverse needs of the hydropower sector, driving operational efficiency, sustainability, and data-driven decision-making.

Table of Contents

List of Figures

List of Tables

List of Acronyms

Acronym	Full description
AI	Artificial Intelligence
DA	Data Analytics bundle
DCAB	Data Collection and Aggregation Bundle
DCS	Data Collection Services bundle
DJB	Data Collection Job
DLT	Distributed Ledger Technology
DoA	Description of Action
DShS	Data Sharing Services bundle
DSL	Data Storage Layer
DSS	Data Security Services bundle
EDIB	External Data Integration Bundle
HPP(s)	Hydro power plant(s)
iAMP	intelligent Asset Management Platform
ICT	Information and Communications Technology
IB	Integration Bundle
iDML	intelligent Data Management Layer
iDOL	Integrated Decision Optimization Layer
ISO	International Organization for Standardization
ML	Machine Learning
Mx	Month x
O&M	Operation and Maintenance
PMM	Predictive Maintenance Module
PMMo	Platform Management Module
PPM	Power Prediction Module
RES	Renewable Energy Sources/Systems
RTDAAB	Real-Time Data Aggregation and Analysis Bundle
SMLCB	Statistical and Machine Learning Core Bundle
Tx.y	Task x.y
UCs	Use Case(s)
WFFM	Weather and Flow Forecasting Module
WPx	Work Package x

1 Introduction

1.1 Description of the deliverable's content and purpose

The present deliverable, titled "D9.1 - Platform architecture designed and released", presents the outcomes of T9.1 and T9.2 activities (pivotal to the initial phases of the iAMP-Hydro project) by month (M) 12 of the project's implementation. T9.1 activities, titled " Definition of Use Cases and Elicitation of user requirements" focused on the definition of the project's Use Cases (UCs) and the elicitation of the iAMP-Hydro project's requirements. This task was critical in gathering and thereafter analysing the requirements for the design and development of the iAMP-Hydro ICT infrastructure consisting of the:

- **iAMP Platform** (iAMP), being the project's key result and which comprises of the:
  - **intelligent Data Management Layer** (iDML)
  - **integrated Decision Optimization Layer** (iDOL).
- **Weather and Flow forecasting Module** (WFFM)
- **Predictive Maintenance Module** (PMM)
- **Power Prediction Module** (PPM)

To ensure that the specifications for the iAMP-Hydro project's ICT components are user-centric and inclusive, synergetic activities were carried out to extract the requirements of these components.
This included the active participation of the IAMP's anticipated (first) end-users, i.e. the project's technical and demonstration partners; while also gathered feedback from external stakeholders who participated in the project's co-development workshops (undertaken in the context of WP13-14).
The input from these activities was then fed directly into T9.2 titled "Detailed architecture design, protocols and interfaces specifications for the Data-driven transformation of the hydro sector"; whose activities are centred around defining the detailed architecture design of the iAMP-Hydro ICT Infrastructure. This includes defining the conceptual architecture, outlining the functionalities and technical specifications of its key components, and designing the individual components.
This dual approach ensures that the architecture not only meets technical standards but also addresses practical user needs of the hydro sector, promoting ease of use, scalability, and security. Thus, D9.1 aims to present a clear, detailed blueprint of the iAMP-Hydro ICT components architecture, illustrating both the broad system's design and the specific details of its components to ensure the effectiveness and adaptability of the iAMP in meeting the evolving demands of the project and the hydro-energy sector in general. Under this context and in accordance with the iAMP-Hydro DoA, D9.1, reports on the design of the iAMP-Hydro ICT components architecture, including, the definition of use cases and user requirements, the conceptual architecture design; the components functional and technical specifications; and the design of individual components.
It shall be noted that any updates (if applicable) on the architecture and/or the intra-platform components will be documented in the D10.1 - 1st release of the iAMP-Hydro iDML,(due in M20), where also the IAM Platform will be described in more detail.

1.2 Structure of the document

To fulfil the objectives of both T9.1 and T9.2 the remainder of this deliverable is structured as follows:

- Chapter 2 presents the methodology followed for the definition of the iAMP-Hydro UCs and the extracted UCs, as defined per Work Package (WP).
- Chapter 3 presents the requirements extraction process, that was followed towards extracting the project's requirements, that in turn will drive the design of the iAMP architecture.
- Chapter 4 presents the conceptual architecture of the iAMP Hydro ICT infrastructure, describing in detail the various components involved and their subcomponents, along with their functional and technical specifications.
- Finally in chapter 6 the conclusion of this deliverable is provided along with the foreseen next steps.

2 iAMP-Hydro Use Cases Definition

In order to comprehensively understand and articulate the diverse applications and benefits of the iAMP-Hydro solution, it is essential to define specific project use cases (UCs), namely the iAMP-Hydro UCs.
A use case is a detailed description of how users interact with a system to achieve a specific goal. It defines the interactions between the user (often referred to as an "actor") and the system. Use cases are particularly valuable in the early stages of software development, as they help capture functional requirements and ensure that all stakeholders have a shared understanding of how the system should behave. Overall, the iAMP-Hydro UCs will serve as practical examples that demonstrate how the proposed solution (i.e. the iAMP-Hydro ICT Infrastructure) can be implemented to address various challenges within the hydropower sector. In essence, the iAMP-Hydro UCs aim to illustrate the potential impacts, benefits, and improvements that the iAMP-Hydro system can bring to hydropower operations, maintenance, and management.

In this direction, this chapter outlines the methodological framework used to identify and define the iAMP-Hydro UCs and provides a detailed template that partners utilised to describe their specific use cases. The goal of using this template was to achieve a uniform expression and understanding of each use case, ensuring clarity and consistency across all contributions.

2.1 iAMP-Hydro Use Cases Template

Towards a uniform expression of the iAMP-Hydro UCs, a detailed template was generated, based on the IEC 62559-1 standard, which defines templates for use cases, actor list and requirements list. The iAMP-Hydro Use Case template (shown in the following table) was then circulated among the partners to be utilised for the definition of their UCs.

Table 3: iAMP-Hydro Use Case Template

Title	**Title of the Use Case**	
ID	*UC_WPno_x*	
Description	*A detailed description of the Use Case (see example provided) without any technical details* • *Why is this Use case needed in iAMP-Hydro, what challenges is trying to solve?* • *What is the result of this use case?*	
Priority	*1 - Very Low, 2 - Low, 3 - Medium, 4 - High 5 - Very High*	
Triggering event	*The event that triggers this UC*	
Preconditions	*Include the key preconditions for this UC*	*Link with other UC(s).*
	Include the key preconditions for this UC	*Link with other UC(s).*
	Include the key preconditions for this UC	*Link with other UC(s).*
Postconditions	*Include the postcondition for this UC.*	
Actors involved	*Define the stakeholders involved in this Use Case.* *(e.g.: data owners, data consumers, Hydro plant Operators, Hydro electromechanical Equipment Vendors, Sensor Technology developers, Energy market stakeholder, etc)*	

2.2 iAMP-Hydro Use Cases definition

The iAMP-Hydro UCs were developed through the collaborative efforts of all consortium partners, who actively participated in a series of structured activities designed to capture the diverse requirements and challenges of the hydropower industry.

The process initiated with the presentation of a draft use case example, using the above template to illustrate the expected context and content. WP Leaders and relevant Task Leaders then provided input specific to their areas of focus. This input was refined through iterative feedback loops, ensuring that the final iAMP-Hydro UCs comprehensively and accurately represented the needs and requirements of the different WPs.

In total, 16 use cases have been defined which include:

- 1 UC for WP2
- 1 UC for WP3
- 2 UCs for WP4
- 3 UCs for WP6
- 6 UCs for WP10
- 4 UCs for WP11

A brief summary of the iAMP-Hydro UCs is presented as follows, followed by their detailed definition (see Tables 4-20), while their mapping to the respective WPs is shown below and in Figure 1:

WP2: Condition Monitoring Sensor Development

UC_2_1: Adaptable condition monitoring sensor suite for enhanced predictive maintenance of hydropower electromechanical equipment
Description: This UC involves the development of a versatile sensor suite designed to monitor the condition of hydropower electromechanical equipment; aiming to enhance predictive maintenance capabilities, allowing for timely interventions and reducing downtime, thus improving the reliability and efficiency of hydropower operations.

WP3: Predictive Maintenance Modelling

UC_3_1: Streamlined failure detection in hydropower plants through automated AI analysis
Description: This UC focuses on leveraging AI technology to streamline the detection of failures in hydropower plants. Automated AI analysis processes operational data to identify patterns and anomalies indicative of potential failures. This proactive approach enables more effective maintenance scheduling and reduces the risk of unexpected equipment breakdowns.

WP4: Ecology Monitoring Sensor Development

- **UC_4_1: Advanced ecological monitoring in hydro power-affected streams**
  Description: This UC is dedicated to the development of advanced monitoring systems for ecological parameters in streams affected by hydropower operations. The goal is to collect detailed and accurate data on ecological health to ensure compliance with environmental regulations and support sustainable hydropower practices.
- **UC_4_2: Identification of possible intervention for ecological performance improvement in hydro plants**
  Description: This UC aims to identify and implement interventions that can enhance the ecological performance of hydropower plants. By analysing ecological data and assessing the impact of hydropower operations, this use case seeks to propose measures that mitigate negative effects on the environment and promote biodiversity

WP6: Flow & available power prediction model development

- **UC_6_1: Improved flow forecasting for enhanced hydropower operations**

Description: This UC involves the development of improved models for forecasting water flow, which are crucial for the efficient operation of hydropower plants. Accurate flow predictions enable better planning and optimization of power generation, leading to increased operational efficiency and reliability.

- **UC_6_2: Data-driven and GIS-supported power availability prediction**
  Description: This UC focuses on creating a data-driven model supported by Geographic Information System (GIS) technologies to predict power availability. By integrating spatial data with historical and real-time information, the model aims to provide more accurate and comprehensive power availability forecasts.
- **UC_6_3: Advanced outflow prediction through AI-driven insights**
  Description: This UC leverages AI to develop advanced outflow prediction models. These models analyse various data sources to provide precise predictions of water outflow, helping to optimize water resource management and hydropower generation.

WP10: iAMP-Hydro Platform – iDML Implementation

- **UC_10_1: Adaptive integration of real-time and batch data via various data collection methods**
  This UC addresses the need for a flexible data integration framework that can handle both real-time and batch data, which by utilising diverse data collection methods, can ensure comprehensive data coverage and support informed decision-making processes in the hydropower industry.
- **UC_10_2: Data enrichment, harmonization, and linking**
  This UC focuses on enhancing the quality and usability of hydro-energy related data through enrichment, harmonization, and linking. It aims to standardise data from multiple sources, making it more coherent and valuable for analysis.
- **UC_10_3: Enhancement of data quality and integrity**
  This UC aims to improve the quality and integrity of data within the iAMP-Hydro platform. By implementing robust data validation and cleansing processes, it ensures that the data used for analysis is accurate and reliable.
- **UC_10_4: Comprehensive governance over data access within an organization**
  This UC involves establishing comprehensive data governance policies to manage access within an organization. It ensures that data is accessible to authorized users while maintaining security and compliance with regulations.
- **UC_10_5: Secure handling of data, mitigating risks of data exposure**
  This UC addresses the need for secure data handling practices to mitigate the risks of data exposure. It involves implementing security measures to protect sensitive information and ensure data privacy.
- **UC_10_6: Efficient querying of data assets and secure acquisition**
  This UC focuses on developing efficient querying mechanisms for data assets, facilitating easy and secure data acquisition. It aims to streamline data access processes, enhancing user experience and productivity.

WP11: Condition Monitoring Sensor Development

- **UC_11_1: Advance analytics insights by integrating proprietary and external data sources**
  This UC aims to generate advanced analytical insights by integrating data from proprietary and external sources; leveraging diverse datasets to provide a holistic view and uncover deeper insights into hydropower operations.
- **UC_11_2: Provision of data analytics baseline models**
  This UC involves the development and provision of baseline analytics models that can be used as a foundation for more complex analyses. These models serve as a starting point for data exploration and decision-making.
- **UC_11_3: Optimized management of hydropower assets**
  This UC focuses on optimizing the management of hydropower assets through data analytics. By analysing operational data, it aims to improve asset performance, extend lifespan, and reduce maintenance costs.

- **UC_11_4: Design and execution of new added-value analytics based on baseline algorithms**
  This UC involves designing and executing new analytics that build upon baseline algorithms. It aims to create added-value analyses that provide actionable insights and drive continuous improvement in hydropower operations.

This structured approach for defining the iAMP-Hydro project's UC, ensures that all aspects of hydropower operations (at least for the project's demonstration sites), from condition monitoring to ecological impact and data analytics, are comprehensively addressed; while the alignment across the project's technical WPs facilitates a cohesive and integrated development process, leading to more effective and innovative solutions than can support the hydropower domain.

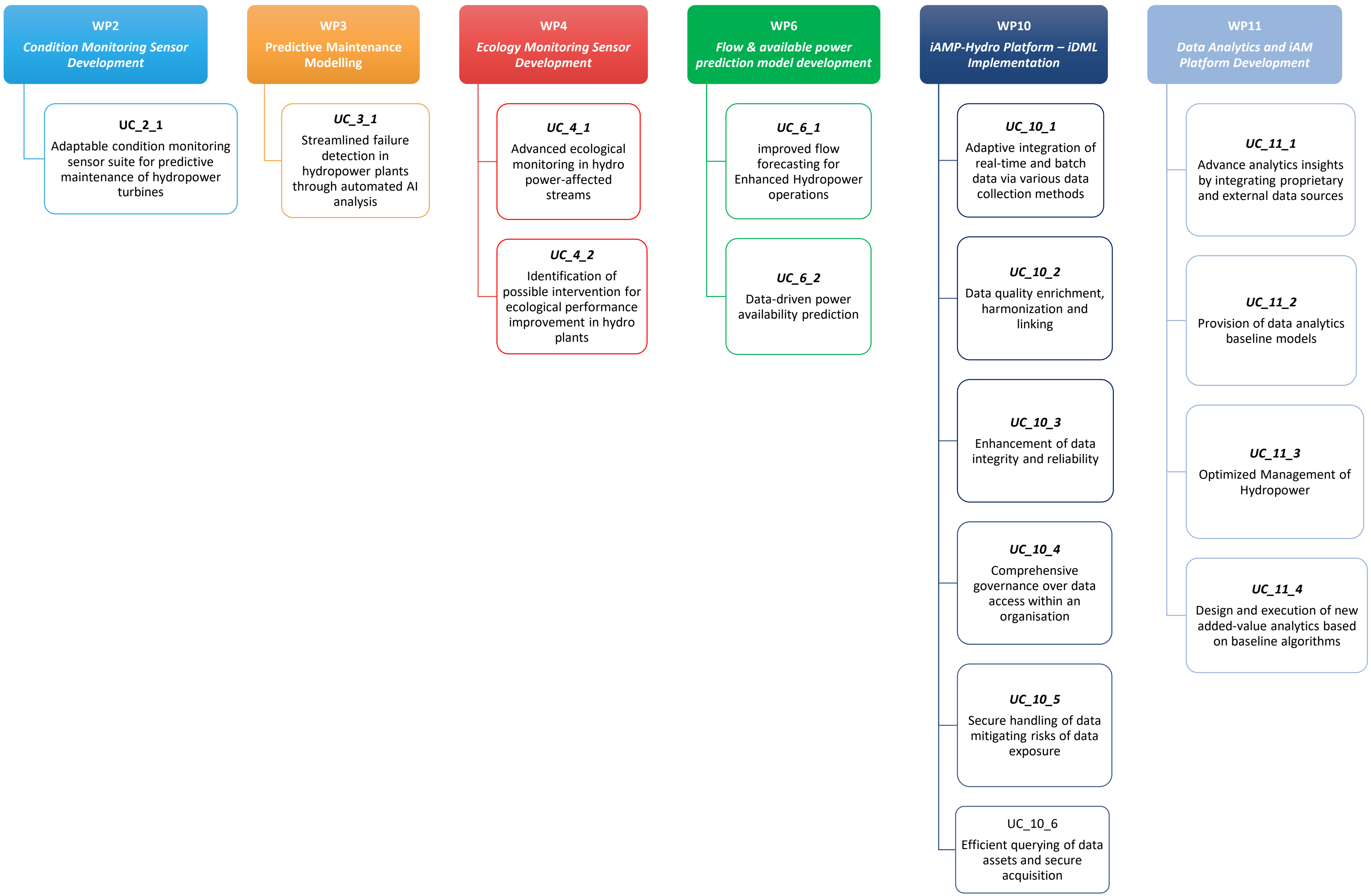

Figure 1 Summary of iAMP-Hydro Use Cases per WP

The detailed descriptions of the Use Case relevant to the **WP2 - Condition Monitoring Sensor Development**, is presented in the following table.

Table 4 iAMP-Hydro UC_2-1

Title	Adaptable condition monitoring sensor suite for predictive maintenance of hydropower turbines	
ID	UC_2_1	
Description	Hydropower has a proven ability in providing not only renewable energy but also delivering this energy at the right time, which is referred to as grid flexibility. The flexibility of hydropower has become more important than in the past for grid operations due to the rise of variable source renewables. Therefore, the application of predictive maintenance not only helps the maintenance aspect, but also supports enhancing output, improving availability, and making investments more profitable. Adopting predictive maintenance into practice for any machine essentially relies on historical data from the same machine or similar kinds of machines. Hydropower electromechanical equipment such as the turbines are forced to operate in off-design conditions to support grid flexibility. This includes frequent transient and part load operations. Such phenomena have several negative repercussions on the functioning state and the remaining useful life of the machine and components. Therefore, efficient condition monitoring with an appropriate set of sensors to measure vibration, acoustic emission, pressure, flow, and speed is crucial for acquiring the necessary data required for predictive maintenance activities.	
Priority	5 - Very High	
Triggering event	New data fed by the sensors into the database.	
Pre-conditions	Set of condition monitoring sensors with specifications appropriate to the type of turbine	N/A
	Set of condition monitoring sensors installed at the demo partners' hydropower plants	N/A
	Availability of robust data collection and management plans.	N/A
Post-conditions	Sensor data is fed into the data management system for further processing and analysis for fault diagnosis and failure prognosis.	
Actors involved	Producer/ Hydro-Plant Operator	

The detailed description of the Use Case relevant to the **WP3 - Predictive Maintenance Modelling**, is presented in the following table.

Table 5 iAMP-Hydro UC_3-1

Title	**Streamlined failure detection in hydropower plants through automated AI analysis**	
ID	UC_3_1	
Description	Hydropower plants are an important contributor to the renewable energy sector. However, due to higher penetration of intermittent renewable sources such as wind and solar, hydro turbines are more regularly operated away from their best efficiency point to fill the demand gaps created. Such fluctuations of load and associated hydrodynamic phenomena imposed on hydro turbines, coupled with their geometrical complexity make them susceptible to different kinds of faults, possibly leading to failures. Faults are characterised by material erosion, cracks, rotor misalignments, etc. due to cavitation, fatigue and wear, resulting in reduced efficiency and power output. Thus, mitigating the risk of turbine failure through predictive maintenance and early fault detection mechanisms is important to ensure the longevity of hydropower electromechanical equipment. Automated AI analysis offers a reliable solution for fault detection in hydropower plants with the integration of sensor data, historical records, and advanced deep-learning algorithms. The goal is to identify some patterns that indicate deviation from the normal working conditions. This will not only enable the plant operators but also the maintenance engineers to address potential issues before they escalate.	
Priority	5 - Very High	
Triggering event	New data fed by the sensors into the database.	
Pre-conditions	Set of condition monitoring sensors with specifications appropriate to the type of turbine	N/A
	Set of condition monitoring sensors installed at the demo partners' hydropower plants	N/A
	Availability of robust data collection and management plans.	N/A
Post-conditions	Sensor data is fed into the data management system for further processing and analysis for fault diagnosis and failure prognosis.	
Actors involved	Producer/ Hydro-Plant Operator	

The detailed descriptions of the Use Cases relevant to the **WP4** - ***Ecology Monitoring Sensor Development***, presented in the following tables.

Table 6 iAMP-Hydro UC_4-1

Title	Advanced ecological monitoring in hydro power-affected streams	
ID	UC_4_1	
Description	In order to meet requirements for sustainable energy production, hydropower needs to have control over physical parameters affecting the biota and have the possibility to react to harmful physical conditions. Physical parameters up-and downstream of the powerplant, needs to be monitored to detect deviations or harmful levels of total dissolved gas supersaturation (TDGS), changes in temperature, and water level (e.g., give low flow or high ramping rate indications in the case of hydropeaking). Chemical parameters such as oxygen and pH need to be incorporated into the system and logged using off-the shelve sensors. Monitoring data needs to be available online and/or locally in real time and delivered via site specific communication protocols (4g, cable or LoRA). Ecological pressures (low water levels or quick changes in water level, critical TDG or temperature values) then need to be linked to operational conditions leading to these values.	
Priority	5 - Very High	
Triggering event	Upload of measured values to database with predefined intervals, or if values outside a predefined values are measured. Power producers need to monitor/improve ecological conditions or monitor/fulfil regulatory requirements.	
Pre-conditions	Site access and possibility for data transfer.	N/A
	Safe place for placing the probes in the water (avoid floods, debris, etc.).	N/A
	Availability of local service personnel for maintenance.	N/A
Post-conditions	Sensor data is fed into the data management system for further processing and analysis for ecological improvements.	
Actors involved	Producer/ Hydro-plant Operator	

Table 7 iAMP-Hydro UC_4-1

Title	**Identification of possible intervention for ecological performance improvement in hydro plants**	
ID	UC_4_2	
Description	In order to meet requirements for sustainable energy production, hydropower needs to know site specific ecological conditions and their response to hydro power activities. Production, flow and biological data needs be collected and basic rules for interventions and more environmentally sustainable operation proposed, to steer the powerplant to reduce the impact on the biota and to avoid harmful conditions (e.g., avoiding certain operation conditions which lead to supersaturation, or increasing downramping time below a certain discharge to avoid stranding of fish).	
Priority	5 - Very High	
Triggering event	Occurrence of environmentally harmful or critical physical conditions.	
Pre-conditions	Availability of time series with critical and non-critical conditions.	N/A
	Possibility to add interventions or rules into the power steering	N/A
	Willingness of the power producer to add interventions	N/A
Post-conditions	Sensor data is fed into the data management system for further processing and analysis, controlling/refining the set rules.	
Actors involved	Producer/ Hydro-Plant Operator	

The detailed descriptions of the Use Cases relevant to the **WP6** - ***Flow & available power prediction model development*** are presented in the following tables.

Table 8 iAMP-Hydro UC_6-1

Title	Improved flow forecasting for Enhanced Hydropower operations	
ID	UC_6_1	
Description	Hydropower relies on natural water flows, subject to seasonal and annual variations, leading to intermittent conditions that affect the consistent provision of reliable power, especially for run of river hydropower plants. Consequently, accurate flow forecasting becomes imperative for effective water management. Different flow forecasting modelling approaches, including AI-based models, physical models or hybrid methods, will be developed for enhanced hydropower operation. The models undergo through a validation process to ensure accuracy and reliability, adapting to new data and operational feedback. Additionally, it provides data necessary for calculating forecasted power availability (UC_6_2).	
Priority	5 – Very High	
Triggering event	Hydro operator's need for decision-making support, to enhance their operations	
Pre-conditions	Historical data on water flow (run of river), reservoir inflow and reservoir levels (reservoir) from hydro operators is readily available.	N/A
	Data (in real-time and/or batch format) available from diverse data sources and of different format.	N/A
Post-conditions	Hydro Operator will be provided with a 7-day water flow / reservoir level prediction. In scenarios where reservoir levels are insufficient, operators will receive instead a medium-term forecast of water level trends to aid in strategic planning.	
Actors involved	River basin authorities, data owners, hydro-plant operators, developers of forecast models.	

Table 9 iAMP-Hydro UC_6_2

Title	**Data-driven power availability prediction**	
ID	UC_6_2	
Description	Flow forecast models (UC_6_1), along with additional data as energy market prices or ecological parameters will be used to provide a sustainable power forecasting model. Prediction of power availability using physical and/or AI-based models provides critical information for planning activities related to the Hydropower plant (HPP) operation including water use, hydropower O&M scheduling, flood prediction, guaranteeing in parallel good ecological conditions for flora and fauna. These models will serve as a foundation for assessing various hydropower production scenarios, considering technical-economic, socio-economic, or ecological factors to ensure the sustainable operation of the HPPs. These models undergo continuous validation to ensure their accuracy and reliability, adapting to new data and operational feedback.	
Priority	5 – Very High	
Triggering event	Hydro operator's need for decision-making support, to enhance their operations	
Pre-conditions	Availability of water flow and reservoir level forecasts	UC_6_1
	Water levels are adequate for the operational requirements of the turbines.	N/A
	No other restrictions of water reserved for other water uses than hydropower (in the case of a multipurpose reservoir) or generated by the reservoir operating rules	N/A
	Data assets (in real-time and batch format) available from diverse data sources and of different format.	N/A
Post-conditions	Hydro Operator will be provided with a 7-day available power prediction.	
Actors involved	River basin authorities, data owners, hydro-plant operators, AI model developers	

The detailed descriptions of the Use Cases s relevant to the ***WP10 - iAMP-Hydro Platform - iDML Implementation***, are presented in the following tables.

Table 10: iAMP-Hydro UC_10-1

Title	**Adaptive integration of real-time and batch data via various data collection methods**	
ID	UC_10_1	
Description	For stakeholders on the hydro-energy sector, it is essential to have access to advanced and highly adaptable systems that facilitate both real-time and batch data collection, in a manner that is transparent and reliable, tailored to their individual organisations' preferences. For the case of batch data, the ingestion process shall be adaptable to the data specificities of the various stakeholders and provide various methods addressing the ingestion of data in batch form (direct file uploads) or via Application Programming Interfaces (APIs) provided by the data providers' systems. In the case of real-time data, the ingestion should be supported by sophisticated streaming data management technologies. All these data ingestion services should be available through a cloud-based environment, utilising secure and trusted data storage.	
Priority	5 - Very High	
Triggering event	Data Providers' willingness to upload data assets onto the data platform for further processing.	
Pre-conditions	Data assets (in real-time and/or batch format) available from diverse data sources and of different format.	N/A
	Properly configured data sources and available interfaces to connect.	N/A
	Availability of an intuitive mechanism enabling data collection via multiple modalities	N/A
Post-conditions	Data assets ingested and available in the platform for further processing based on the data providers' requirements.	
Actors involved	All stakeholders of the hydro-energy domain acting as Data Providers.	

Table 11 iAMP-Hydro UC_10-2

Title	Data enrichment, harmonization and linking	
ID	UC_10_2	
Description	Data assets uploaded onto a software platform inherently possess value; however, their value significantly increases when these are enriched with additional data at both semantic and payload levels. This enrichment enhances the overall data quality and insights that can derive from it, while simplifies the process of identifying other data assets that could be synergistically combined to create more complex and value-added derivative data assets. This approach not only boosts the discoverability and interoperability of the data, but also uncovers potential relationships with other assets. Moreover, data users should have the capability to manipulate and merge these data assets, incorporating data from various sources to create new data assets filled with richer, more valuable insights. However, data assets utilised within the hydro-energy domain, are typically of different format and derive from diverse and heterogeneous sources (e.g. legacy systems, operational systems, APIs, historical records, statistics, sensors/IoT devices, weather data, energy markets, open data sources, etc.). As such, it is vital to provide the capability to stakeholders to integrate and enhance their data assets from these multiple sources simultaneous, through the delivery of an integrated mechanism that can assist them in the representation and analysis of data from these varied sources. Consequently, this mechanism should offer a process that involves hydro-energy domain stakeholders directly, allowing them to participate effectively and effortlessly; while enabling them to precisely define the context and semantics of their data, ensuring their data can be fully leveraged and utilised.	
Priority	5 - Very High	
Triggering event	New data assets uploaded on to the platform.	
Pre-conditions	Availability of hydro-energy data assets from various diverse data sources and of different formats	UC_10_1
	Availability of a versatile mechanism facilitating the ingestion of real-time or batch energy data assets	UC_10_1
	Availability of a versatile and highly customisable mechanism facilitating the representation and semantic interpretation and linking of the ingested data assets	UC_10_1
Post-conditions	Diverse data assets are accurately ingested in the platform represented with the correct context and semantics, and available for further pre-processing	
Actors involved	All stakeholders of the hydro-energy domain acting as Data Providers	

Table 12 iAMP-Hydro UC_10_3

Title	Enhancement of data quality and Integrity	
ID	UC_10_3	
Description	In the hydro-energy domain (also applicable across various sectors), ensuring the integrity and quality of the data utilised is critical for accurate analysis, decision-making, and operational efficiency. Data assets typically originate from various sources (legacy systems, operational systems, historical records, sensors, etc.). These diverse sources can introduce inconsistencies, errors, and potential data corruption. To address these challenges, it is essential to implement robust mechanisms for data validation, error detection, and correction. Thus, priority shall be given in enhancing the integrity and quality of data assets by deploying advanced data curation techniques and robust methods; involving continuous monitoring of data quality, identifying and rectifying discrepancies, and ensuring that data remains consistent and accurate over time. By improving data integrity and quality, stakeholders can trust the data they use for critical applications, leading to better decision-making and operational outcomes. Such mechanisms should enable stakeholders to define validation rules, automate data checks and generate alerts for any detected issues, ensuring that the integrated data is reliable and ready for further analysis and use.	
Priority	5 - Very High	
Triggering event	Uploading of new data asset on the platform	
Pre-conditions	Availability of hydro-energy data assets from various diverse data sources and of different formats	UC_10_1
	Availability of a versatile mechanism facilitating the ingestion of real-time or batch energy data assets	UC_10_1
Post-conditions	The integrity and quality of the data assets is considerably enhanced prior to their storage in the data platform.	
Actors involved	All stakeholders of the hydro-energy domain acting as Data Providers	

Table 13 iAMP-Hydro UC_10_4

Title	Comprehensive governance over data access within an organisation	
ID	UC_10_4	
Description	In any organisation, particularly within the hydro-energy domain, managing who has access to data and ensuring that data access is governed by robust policies and procedures is critical for maintaining data security, privacy, and compliance. To this end it is of utmost importance that appropriate mechanisms are in place to establish and enforce a comprehensive data access governance framework/mechanism within data platforms, that can ensure comprehensive governance over data access within organisations. Such mechanisms shall facilitate the definition of clear policies for data access rights and role-based access control; by doing so, the organisation can ensure that only authorized personnel have access to confidential data, monitor how data is being used, and respond quickly to any unauthorized access or anomalies. The mechanisms should allow stakeholders to define and manage access policies, assign access rights based on personnel roles and responsibilities, and track all data access activities.	
Priority	5 - Very High	
Triggering event	New data asset ingested into the platform or changes in its access policies	
Pre-conditions	Availability of hydro-energy data assets from various diverse data sources and in different formats	UC_10_1
	Implementation of a versatile and highly customizable mechanism for managing data access rights and roles	N/A
Post-conditions	• Data access rights are managed effectively, ensuring that only authorized personnel have access to sensitive data • Organisations can respond promptly to any unauthorized access or anomalies, maintaining data security and privacy	
Actors involved	All stakeholders of the hydro-energy domain acting as Data Providers, Data Consumers and organisations' IT security personnel	

Table 14 iAMP-Hydro UC_10_5

Title	**Secure handling of data, mitigating risks of data exposure**	
ID	UC_10_5	
Description	In the hydro-energy domain, data assets originate from various sources, (legacy systems, operational systems, historical records, sensors/IoT devices, etc.); given the sensitivity and critical nature of these data assets, it is imperative to ensure their secure handling within data platforms to mitigate risks associated with data exposure, such as unauthorized access, data breaches, and compliance violations. In this context, robust security measures need to be put in place to protect data from potential threats and vulnerabilities. Such mechanisms should enable stakeholders to implement and enforce security policies on their data assets by effectively isolating and protecting confidential and sensitive information prior to any further processing, thus protecting their business interests, removing concerns related to data security, increasing trust in data sharing and ensuring compliance with data protection regulations	
Priority	5 - Very High	
Triggering event	New data asset ingested into the platform or updates to data asset's access policies	
Pre-conditions	Availability of hydro-energy data assets from various diverse data sources and in different formats.	UC_10_1
	Data assets uploaded onto the platform.	UC_10_2
	Availability of a robust mechanism for encrypting data at rest and in transit.	N/A
Post-conditions	• Data within the is securely handled, with strong access controls in place. • Any security incidents are promptly addressed, mitigating risks of data exposure and breaches.	
Actors involved	All stakeholders of the hydro-energy domain acting as Data Providers, and organisations' IT security and compliance personnel.	

Table 15 iAMP-Hydro UC_10_6

Title	Efficient querying of data assets and secure acquisition	
ID	UC_10_6	
Description	The ability to efficiently query and securely access data assets is crucial for enabling timely and informed decision-making and in the case of hydropower plants this is more than evident. To maximise the utility of the diverse datasets utilised from hydro/energy sector stakeholders, a robust data exploration mechanism is needed that will allow the quick and efficient search and discovery of relevant data shared by the variety of stakeholders involved in the hydropower value chain. In this context, focus shall be given on implementing highly efficient querying mechanisms that can support complex queries across large datasets, ensuring their quick data retrieval, upon the agreement of both parties and the signing of a smart data contract. In this way stakeholders can extend their data portfolio by obtaining access to external datasets from diverse sources to enrich analysis capabilities, improve insights, and drive innovation. Additionally, special emphasis shall be given to the establishment of secure data sharing processes to protect sensitive information and ensure compliance with data protection regulations. This includes secure authentication methods and enforcement of access controls to data to restrict unauthorized access. Furthermore, it is imperative that data querying mechanisms and data discovery adhere strictly to the established data access policies. These policies ensure that no unauthorized data consumer can access data they are not entitled to, maintaining the integrity and confidentiality of sensitive information. Both the querying and smart contracting mechanisms need to be delivered through user-friendly interfaces, combining efficient querying capabilities with secure data sharing protocols, and immutable smart data contract mechanisms, based on Distributed Ledger Technologies (DLT) so that stakeholders can access the data they need swiftly and safely, enhancing their analytical capabilities and operational efficiency.	
Priority	5 - Very High	
Triggering event	Requirement for accessing specific data assets for utilisation in analysis or operational purposes	
Pre-conditions	Data assets available and uploaded onto the platform	UC10_1, UC10_2, UC10_3
	Implementation of a versatile and highly efficient querying engine.	N/A
	Deployment of smart contract engine, based on DLT-enabled infrastructure	UC10_5, UC11_1
	Deployment of secure data acquisition mechanisms, including authentication, encryption, and access control.	N/A
Post-conditions	• Data assets are efficiently queried and securely acquired, ensuring quick and safe access • Stakeholders acquire the data needed through the signature of smart data contract, backed by a DLT-enabled infrastructure • Stakeholders can perform complex data retrievals with minimal delay and maximum security • Enhanced data accessibility and security support better decision-making and operational processes	
Actors involved	All stakeholders of the hydro-energy domain acting as Data Consumers and Data Providers	

The detailed descriptions of the Use Cases relevant to the **WP11** - ***Data Analytics and iAM Platform Development***, are presented in the following tables.

Table 16 iAMP-Hydro UC_11_1

Title	**Advance analytics insights by integrating proprietary and external data sources**	
ID	UC_11_1	
Description	In the hydro-energy sector, leveraging advanced analytics to gain valuable insights is critical for optimising operations, enhancing decision-making, and driving innovation. This requires the integration of both proprietary data (such as operational data, historical performance data, and sensor data) and external data sources (such as weather data, market trends, and public datasets). By combining these diverse data sources, stakeholders can uncover deeper insights, identify patterns, and make more informed decisions. To this end, a unified approach and robust technological tools are required to enable the seamless integration of proprietary and external data sources into advanced analytics tools and platforms. The latter shall support advanced analytics techniques, including machine learning (ML) and predictive analytics, as well as data visualisation capabilities. This will allow stakeholders to not only perform sophisticated analyses but also to visualize insights clearly, providing contextual understanding and facilitating more informed decision-making. By integrating various data sources and combing proprietary data with external, stakeholders can develop comprehensive analytics models that provide actionable insights, optimise resource utilization, and can improve hydropower plants' overall operational efficiency.	
Priority	5 - Very High	
Triggering event	Availability of new proprietary or external data sources, or a need for enhanced analytics capabilities.	
Pre-conditions	Availability of diverse proprietary and external data sources in various formats and uploaded in the platform.	N/A
	Availability of an intuitive mechanism enabling data collection via multiple modalities.	UC_10_1
	Availability of should also provide tools for data cleansing, transformation, and enrichment to enhance data quality and facilitate accurate analysis.	UC_10_2
	Provision of an advanced analytics engine that supports data preparation, ML and predictive analytics, and data visualisation functionalities.	N/A
Post-conditions	• Data from proprietary and external sources are seamlessly integrated into the platform. • Stakeholders can derive advanced analytics insights to support decision-making and hydro power plants' operational optimisation.	
Actors involved	All stakeholders in the hydro-energy sector, including Data Analysts, Data Scientists, and Decision Makers	

Table 17 iAMP-Hydro UC_11_2

Title	Provision of data analytics baseline models	
ID	UC_11_2	
Description	Optimization of operational efficiency across hydropower stakeholders needs to rely in data-driven insights delivered through the utilization of AI models. Considering the lack of data science expertise within the hydropower sector, it is imperative that AI models are provided in an easy-to-use manner, stepping on pre-trained pipelines that can be easily adapted to the unique context of each hydropower plant. To this end comprehensive data analytics pipelines need to be created trained and offered as baseline artefacts, in the form of pre-trained models addressing specific needs for predictive maintenance, biodiversity assessment and monitoring, and weather/flow forecasting. Such models can provide a sound basis for further enhancing the efficiency, reliability, and sustainability of hydro-energy operations by providing advanced analytical insights and supporting data-driven optimization processes in hydropower plants, while ensuring equipment reliability, minimizing environmental impacts, and improving overall decision-making processes for all stakeholders involved.	
Priority	5 - Very High	
Triggering event	Need to derive data analytics insights	
Pre-conditions		
	Availability of diverse proprietary and external data sources in various formats and uploaded in the platform.	N/A
	Delivery of ready-to -use data analytics models	N/A
	Availability of an interactive visualisation tool for insights presentation	UC_11_1
Post-conditions	• Successful creation and integration of data analytics models • Established analytics pipelines for predictive maintenance, biodiversity monitoring, and weather/flow forecasting.	
Actors involved	All stakeholders in the hydro-energy sector, including Data Analysts, Data Scientists, and Decision Makers	

Table 18 iAMP-Hydro UC_11_3

Title	Optimized Management for Hydropower Assets	
ID	UC_11_3	
Description	Hydropower plays a critical role in the energy transition due to its inherent flexibility in generation and substantial energy storage capabilities. As we move towards a more sustainable energy mix, the ability of hydropower plants to quickly adjust output to meet fluctuating demand makes them indispensable. Despite their potential, many hydropower assets operate with outdated technologies and methodologies. Modernising these facilities is crucial for enhancing their efficiency, reliability, and overall contribution to the energy grid. To address this challenge development of an advanced optimization tool tailored for hydropower assets would be highly beneficial. This tool will use technologies such as machine learning, the Internet of Things (IoT), and decision support systems to create an optimised management framework. The algorithm will integrate predictive maintenance, biodiversity considerations, and weather and flow forecasting AI data analytics developed in various work packages (WPs 2-7). Additionally, it will incorporate external data related to other renewable energy sources (RES), predicted component lifetimes, maintenance costs, electricity market dynamics, and environmental constraints to determine the optimal operating strategies and maintenance schedules. The aim is to develop advanced optimization algorithms and decision support to balance multiple objectives, such as maximising energy production, minimising operational costs, and ensuring regulatory compliance. Overall, the Optimised Management for Hydropower Assets use case aims to develop a holistic solution that combines predictive maintenance, asset management optimization, operational planning, and environmental considerations.	
Priority	5 - Very High	
Triggering event	Availability of new proprietary or external data sources, or a need for enhanced analytics capabilities.	
Pre-conditions	Hydropower plant is operational and equipped with necessary instrumentation and data acquisition systems	UC_2_1
	Availability of AI analysis solution for fault detection in hydropower plants	UC_3_1
	Data of physical and chemical parameters affecting the ecological conditions in hydropower-affected streams.	UC_4_1
	Availability of Weather/flow forecasting models	UC_6_3
	Availability of historical data on asset performance, maintenance records, and operational parameters.	N/A
	Availability of electricity market data sources.	N/A
	Provision of an advanced analytics engine that supports data preparation, ML and predictive analytics, and data visualisation functionalities.	N/A
Post-conditions	• Predictive maintenance schedules for hydropower assets are generated based on real-time condition monitoring data and forecasted component health • Asset management and lifecycle plans are optimized, considering factors like predicted component lifetimes, maintenance costs, and operational constraints	
Actors involved	All stakeholders in the hydro-energy sector, including Data Analysts, Data Scientists, and Decision Makers	

Table 19 iAMP-Hydro UC_11_4

Title	Design and execution of new added-value analytics based on baseline algorithms	
ID	UC_11_4	
Description	Developing new and improved energy services based on data analysis is a complex and effort-intensive endeavour. It requires performing various data manipulation processes to aggregate the necessary data and transform it into an appropriate format for the core analysis processes. Additionally, the analytics processes themselves must be iteratively applied, assessed, and refined until the results are deemed acceptable, successful, or significant. These processes can be extremely time-consuming and often become demotivating, especially during the experimentation and validation phases of a new service idea. To address these challenges and streamline the development process, a modular and reusable approach to the analytics workflow can be highly beneficial. By defining and providing out-of-the-box access to individual steps of the analytics process, from simple data manipulation functions to baseline algorithms as a service, data scientists can seamlessly combine and integrate these components into the design, development, and deployment of new services. This approach not only increases efficiency but also fosters consistency, reproducibility, and maintainability of the analytics workflows, ultimately leading to more robust and reliable energy services.	
Priority	4 – High	
Triggering event	None	
Pre-conditions	Baseline algorithms have been developed as reusable services and can be used.	N/A
	These baseline algorithms are available for integration and deployment as part of the modular analytics framework.	
Post-conditions	A new analytics process has been designed by combining and orchestrating the reusable data manipulation functions and baseline algorithms from the modular framework.	
Actors involved	All stakeholders in the hydro-energy sector, including Data Analysts, Data Scientists, and Decision Makers	

3 iAMP-Hydro Requirements definition and analysis

3.1 Requirements extraction methodology

Having defined the project's UCs, the next steps entailed the extraction of the requirements stemming from the defined UCs. To ensure the effective development and implementation of the iAM Platform, a structured methodology was followed for extracting the project's requirements. This methodology employed a consistent syntax to define each requirement, enhancing clarity and uniformity. The requirements are categorised by priority to guide the development plan effectively, using specific linguistic parameters to denote their level of necessity.

For the formulation/drafting of the iAMP-Hydro requirements a coherent way was followed by all partners, towards expressing each requirement in the form: [Subject] [Action] [Object] [Value] [Constraint] [Condition] which is further explained below:

- Subject: The system or component being developed, i.e., iAMP-Hydro.
- Action: The specific function or task the system will perform, denoted by a verb.
- Object: The target of the action or what is being acted upon.
- Value: The benefit or purpose of the requirement.
- Constraint: Any limitations or conditions that must be met.
- Condition: Circumstances under which the requirement applies.

As can be seen from the requirement collection template, a priority is also given, while a set of linguistic parameters were followed for their structure. The following guidance was provided to the partners for the description of the

Linguistic Parameters Guidance:

Priority 1-2:Low priority - Use the verb "*may*" for optional suggestions.

Priority 3-4: Medium priority - Use the verb "*should*" for advisable features.

Priority 5: Hight priority - Use the verb "*shall*" for mandatory requirements.

It shall be noted that the requirements prioritisation is based on the requirement's importance and urgency to the project's needs, driving the development timeline accordingly.

Finally, once all requirements were collected, work focused on their validation and verification; ensuring that each requirement is specific, measurable, and testable. In this direction partners were asked to avoid vague terms when defining their requirements and omit comparisons, and incomplete statements to eliminate potential loopholes.
By adhering to this methodology, the iAMP-Hydro project maintained a clear and structured approach to requirements extraction, ensuring that all stakeholders' needs are accurately captured and addressed in the system's development.

3.2 IAMP-Hydro requirements

Following the above methodology the iAMP-Hydro requirements, their priority and their relevance to the project's UCs are provided in the following table.

Table 20 iAMP-Hydro Requirements

Requirement _ID	Description	Priority	Relevant UC(s)
Req_001	iAMP-Hydro shall assist the hydro operators in deciding the appropriate type, number and location of condition monitoring sensors based on the type of machine and other technical characteristics of the plant.	5	UC_2_1
Req_002	iAMP-Hydro shall facilitate hydro operators in data collection and management obtained from the sensors.	5	UC_2_1, UC_3_1
Req_003	iAMP-Hydro should advice hydro operators to act proactively regarding sensor calibrations at regular intervals.	3-4	UC_2_1,
Req_004	iAMP-Hydro may provide instructions for safe management and maintenance of the condition monitoring equipment.	1-2	UC_2_1
Req_005	iAMP-Hydro should classify potential faults based on turbine type and its operational parameters.	3-4	UC_3_1
Req_006	iAMP-Hydro shall indicate the signs of incipient faults, identify the cause (cavitation, fatigue, or wear), if possible, in turbines using the condition monitoring data.	5	UC_3_1
Req_007	iAMP-Hydro shall utilise machine learning algorithms to provide information on the remain useful life of the turbine based on the estimated time to impending failure.	5	UC_3_1
Req_008	iAMP-Hydro should trigger alarms to take maintenance actions.	3-4	UC_3_1
Req_009	iAMP-Hydro shall assist the hydro operators in deciding the appropriate type, number and location of environmental sensors based on the plant design.	5	UC_4_1
Req_010	iAMP-Hydro shall assist the hydro operators in the collection and management of the signal data obtained from the sensors for further analysis.	5	UC_4_1 UC_4_2
Req_011	iAMP-Hydro shall assist the hydro operators in defining rules and threshold values in production, flow and biological parameters, to steer the powerplant to reduce the impact on the biota and to avoid harmful conditions. Resulting in a more environmentally sustainable operation.	5	UC_4_2
Req_012	iAMP-Hydro shall forecast water flow and reservoir levels for a period of 7 days, to assist hydro operators in decision-making processes.	*5*	UC_6_1
Req_013	iAMP-Hydro shall provide output necessary for calculating the forecasted power availability based on water management data.	5	UC6_1
Req_014	iAMP-Hydro shall offer a medium-term forecast of water level trends when current reservoir levels are not sufficient for optimal operation.	*5*	UC_6_1
Req_015	iAMP-Hydro shall undertake data validation tasks on the water flow / water level forecasting before models are deployed.	*5*	UC_6_1 UC_10_1
Req_016	iAMP-Hydro shall predict available power for the next 7 days using physical or AI-based models.	*5*	UC_6_2
Req_017	iAMP-Hydro shall ensure that data assets are available from diverse sources in both real-time and batch formats.	*5*	UC_6_2

Req_018	iAMP-Hydro should maintain a continuous update process for AI models to adapt to changing conditions.	*5*	UC_6_2
Req_019	iAMP-Hydro shall undertake data validation tasks on the available power prediction before models are deployed. This validation task shall be done against real data from each of the validation sites.	*5*	UC_6_2
Req_020	iAMP-Hydro shall store operational data from hydro plants, including water flow, reservoir levels, and energy production, ensuring a comprehensive dataset for model training and validation.	*5*	UC_10_1 UC_6_2
Req_021	iAMP-Hydro shall deliver an intuitive UI enabling users to upload their data assets through multiple modalities.	5	UC_10_1
Req_022	iAMP-Hydro shall enable users to provide samples of their organisation's data assets	5	UC_10_1
Req_023	iAMP-Hydro shall enable users to upload his/her organisation's data assets as single and /or batch files.	5	UC_10_1
Req_024	iAMP-Hydro shall enable data providers to periodically upload his/her organisation's data assets through his/her own organisation's APIs.	5	UC_10_1
Req_025	iAMP-Hydro shall enable data providers to upload his/her organisation's data assets utilising an API offered by the iAMP.	*5*	UC_10_1
Req_026	iAMP-Hydro shall enable users to upload data files of different formats (i.e. json, csv, xml, etc.)	*5*	UC_10_1
Req_027	iAMP-Hydro shall enable users to schedule the future retrieval of data assets from the APIs exposed by their organisation's system.	*5*	UC_10_1
Req_028	iAMP-Hydro shall enable users to create composite data assets by integrating with existing data assets that are already stored in the platform.	5	UC_10_1
Req_029	iAMP-Hydro shall enable users to see a preview of a composite data asset prior to creating it.	5	UC_10_1
Req_030	iAMP-Hydro shall enable users to specify the retrieval schedule of their data collection processes.	5	UC_10_1
Req_031	iAMP-Hydro shall enable users to set-up multiple and concurrent data collection processes.	5	UC_10_1
Req_032	iAMP-Hydro shall harmonise and store all ingested data assets according to a common information data model.	*5*	UC_10_2
Req_033	iAMP-Hydro shall establish a common information model that considers current standards and models pertinent to the hydropower energy sector.	*5*	UC_10_2
Req_034	iAMP-Hydro shall offer a user-friendly interface where data owners during the initial data collection stage, can map (interrelate) their data assets to the domain concepts and attributes of iAMP's common information model in a semi-automated manner.	*5*	UC_10_2
Req_035	iAMP-Hydro shall provide functionality for data owners during the initial data collection stage, to manually adjust the mappings (interrelation) between their data asset's attributes and those of the iAMP's Common Information Model.	*5*	UC_10_2
Req_036	iAMP-Hydro shall be capable of executing all necessary data modifications/transformations related to a data asset's attributes/fields (e.g. measurement units' conversion, date and	*5*	UC_10_2

	time format and time zones transformation, , etc.) adhering to the specifications of the iAMP’s Common Information Model.		
Req_037	iAMP-Hydro shall offer a user-friendly interface allowing data owners during the initial data collection stage, to specify rules for data transformation that should be applied to the attributes of their data asset.	*5*	UC_10_2
Req_038	iAMP-Hydro shall undertake cleansing tasks on the data assets of data owners before these are stored.	*5*	UC_10_3
Req_039	iAMP-Hydro shall execute data cleansing operations on the users’ data assets, based on their data cleaning configurations.	*5*	UC_10_3
Req_040	iAMP-Hydro shall undertake data validation tasks on the data assets of data owners before these are stored.	*5*	UC_10_2 UC_10_3
Req_041	iAMP-Hydro shall users to define their preferred data transformation rules that are to be performed on their data prior to being stored.	*5*	UC_10_3
Req_042	iAMP-Hydro shall undertake data completion operations on the data assets, prior to being stored.	*5*	UC_10_3
Req_043	iAMP-Hydro shall provide organisation-based access to its users.	*5*	UC_10_4
Req_044	iAMP-Hydro shall enable data providers to define the licensing details of their owned data asset.	*5*	UC_10_4
Req_045	iAMP-Hydro shall enable data providers to select a licence from a of predefined list.	*5*	UC_10_4
Req_046	iAMP-Hydro should enable data providers to define their custom data license for their data asset.	*4*	UC_10_4
Req_047	iAMP-Hydro shall enable data providers to define their preferred data access policies of his/her organization's data assets.	*5*	UC_10_4
Req_048	iAMP-Hydro shall allow data providers to upload data that are envisioned as confidential and not to be shared with others.	*5*	UC_10_4 UC_10_5
Req_049	iAMP-Hydro shall be capable to articulate and impose a data access control assessment based on the associated data access policies.	*5*	UC_10_5
Req_050	iAMP-Hydro shall enable data providers to define and select multiple data access policies on the data assets owned by their organisations.	*5*	UC_10_4 UC_10_5
Req_051	iAMP-Hydro shall be capable of storing unencrypted data assets in a secure data storage in its cloud infrastructure.	*5*	UC_10_4 UC_10_5
Req_052	iAMP-Hydro shall enable data providers to update/modify the data access policies of each data asset owned by their organisation.	*5*	UC_10_5
Req_053	iAMP-Hydro shall enable data providers to update/modify licensing details of each data asset owned by their organisation.	*5*	UC_10_5
Req_054	iAMP-Hydro shall allow the definition of data access policies based on the contextual information of the requested asset's attributes.	*5*	UC_10_5
Req_055	iAMP-Hydro shall enable the definition of data access policies based on the requested asset's attributes.	*5*	UC_10_5
Req_056	iAMP-Hydro should allow data providers to upload private data that can be shared with data consumer only under specific circumstances.	*4*	UC_10_5

Req_057	iAMP-Hydro shall enable data providers to upload data that are intended to be publicly available to all data consumers.	5	UC_10_5
Req_058	iAMP-Hydro shall deliver a search query interface with intuitive UI enabling users to search and identify data assets of interest.	*5*	UC_10_6
Req_059	iAMP-Hydro shall enable users to search for available assets in the platform, based on the data assets' contents and access policies.	*5*	UC_10_5 UC_10_6
Req_060	iAMP-Hydro shall enable users to search for available assets in the platform, setting filters based on the data assets' metadata.	*5*	UC_10_6
Req_061	iAMP-Hydro should enable users to search for available assets by setting filters based on the domains these data assets are relevant.	*5*	UC_10_6
Req_062	iAMP-Hydro may enable users to search for available assets by setting filters based on the attributes/fields present in the data assets.	*3*	UC_10_6
Req_063	iAMP-Hydro may deliver suggestions to users of additional datasets (available in the platform) based on their search queries filters.	*3*	UC_10_6
Req_064	iAMP-Hydro may deliver suggestions of other registered organisations that have data assets relevant to their performed search query.	*3*	UC_10_6
Req_065	iAMP-Hydro shall enable users to browse through the assets available to be obtained by their organisation.	*5*	UC_10_6
Req_066	iAMP-Hydro shall enable users to browse through the data assets obtained by their organisation, in accordance with the respective data asset's contracts terms.	*5*	UC_10_6
Req_067	iAMP-Hydro shall deliver an intuitive UI, enable data providers and data consumers to exchange data assets based on smart data contract mechanism.	*5*	UC_10_6
Req_068	iAMP-Hydro shall enable users to define and edit the terms of a smart data contract, before its signature from both parties.	*5*	UC_10_6
Req_069	iAMP-Hydro shall enable users to approve or reject changes on the terms of an asset contract in which their organisation is involved, prior to signing the smart data contract.	5	UC_10_6
Req_070	iAMP-Hydro should enable data providers to replace their organisation's private data assets under certain circumstances, always adhering to the terms of the respective smart data contract.	4	UC_10_6
Req_071	iAMP-Hydro shall enable users to withdraw from a negotiation process over a data asset, prior to both parties signing the smart data contract.	5	UC_10_6
Req_072	iAMP-Hydro shall enable only the authorised representatives of organisations to be involved in a data sharing transaction and sign the respective smart data contract.	5	UC_10_6
Req_073	iAMP-Hydro shall save all the data transaction operations and signed smart data contracts in an immutable list.	5	UC_10_6
Req_074	iAMP-Hydro shall enable the authorised representatives of organisations involved in a smart data contract to download the signed smart data contract as a file.	5	UC_10_6
Req_075	iAMP-Hydro shall ensure that all data sharing services adhere to the terms defined in the relevant signed smart data contract.	5	UC_10_6

Req_076	iAMP-Hydro shall ensure the transparency of all data sharing processes, adhering to the applicable data asset's access policies.	5	UC_10_6
Req_077	iAMP-Hydro shall enable users to expose the results of a data analytics workflow to be consumed through an API.	5	UC_10_6
Req_078	iAMP-Hydro should allow a user to upload an available data analytics workflow through an API.	4	UC_10_6
Req_079	iAMP-Hydro shall guarantee the necessary cloud resources are available for its services' efficient operation.	5	UC_10_1 UC_10_2 UC_10_3 UC_10_4 UC_10_5 UC_10_6
Req_080	iAMP-Hydro shall deliver an intuitive UI enabling users to design and execute data analytics workflows	*5*	UC_11_1
Req_081	iAMP-Hydro shall enable users to combine different data manipulation-preparation and data analysis methods in a pipeline to be executed sequentially as a service.	*5*	UC_11_1
Req_082	iAMP-Hydro shall enable users to perform a data analytics workflow immediately or set a schedule for its execution.	*5*	UC_11_1
Req_083	iAMP-Hydro shall enable users to compute aggregations over a data asset (they own or have legitimate acquired).	*5*	UC_11_1
Req_084	iAMP-Hydro shall enable users to choose from a list of ML/DL algorithms and apply them to a data asset (they own or have legitimate acquired).	*5*	UC_11_1
Req_085	iAMP-Hydro shall enable users to configure the execution parameters of the available ML algorithms so as to be used in their data analytics pipeline.	*5*	UC_11_1
Req_086	iAMP-Hydro shall enable users to extract features from a data asset to be used as input by a ML algorithm.	*5*	UC_11_1
Req_087	iAMP-Hydro shall enable users to extract features from a data asset to be used as input in a data analytics pipeline.	*5*	UC_11_1
Req_088	iAMP-Hydro shall enable users to combine data from multiple data assets (they own or have legitimate acquired).	*5*	UC_11_1
Req_089	iAMP-Hydro shall safeguard the ML/DL models utilised against intended biases that could manipulate the training dataset and underlying algorithms.	*5*	UC_11_1
Req_090	iAMP-Hydro shall enable the detection of biases within the training datasets and underlying algorithms.	*5*	UC_11_1
Req_091	iAMP-Hydro should provide mechanisms to correct detected biases before model training and deployment.	*4*	UC_11_1
Req_092	iAMP-Hydro should continuously monitor the ML/DL models for any bias introduction during model updates or data changes.	*4*	UC_11_1
Req_093	iAMP-Hydro shall establish and track baseline performance metrics (e.g., accuracy, F1 score, etc.), to assess and monitor the performance of its ML/DL models.	*5*	UC_11_1
Req_094	iAMP-Hydro shall ensure the transparency of its ML/DL models' by assessing model transparency aspects (e.g., decomposability and algorithmic transparency)	*5*	UC_11_1

Req_095	iAMP-Hydro shall ensure the transparency of its ML/DL models' by assessing transparency aspects evaluated by target audiences (e.g., satisfaction, fidelity, completeness, ambiguity).	*5*	UC_11_1
Req_096	iAMP-Hydro should deliver real-time alerts and logs when biases are detected, providing real-time feedback to users.	*4*	UC_11_1
Req_097	iAMP-Hydro shall enable users to create reports on data owned by their organisation.	*5*	UC_11_1
Req_098	iAMP-Hydro shall enable users to save a report.	*5*	UC_11_1
Req_099	iAMP-Hydro shall enable users to save the results of a data analytics pipeline and export them as downloadable file.	*5*	UC_11_1
Req_100	iAMP-Hydro shall enable users to visualise the results of a data analytics pipeline they created.	*5*	UC_11_1
Req_101	iAMP-Hydro shall enable users to generate visualisations on the data owned/obtained by their organisation.	*5*	UC_11_1
Req_102	iAMP-Hydro shall enable users to save a visualisation they created.	*5*	UC_11_1
Req_103	iAMP-Hydro shall enable users to access other visualisations owned (or legitimate acquired) by their organisation.	*5*	UC_11_1
Req_104	iAMP-Hydro shall deliver pre-trained AI models that are easy to use and adapt to the unique contexts of individual hydropower plants.	*5*	UC_11_2
Req_105	iAMP-Hydro shall offer baseline AI algorithms in the form of pre-trained models specifically designed for Predictive maintenance.	*5*	UC_11_2
Req_106	iAMP-Hydro shall offer Predictive Maintenance models that can be customised to accommodate different equipment types and operational contexts.	*5*	UC_11_2
Req_107	iAMP-Hydro shall offer baseline AI algorithms in the form of pre-trained models specifically designed for Biodiversity assessment and monitoring.	*5*	UC_11_1 UC_11_2
Req_108	iAMP-Hydro shall offer baseline AI algorithms in the form of pre-trained models specifically designed for Weather and flow forecasting.	*5*	UC_11_2
Req_109	iAMP-Hydro shall provide an interactive visualisation tool for presenting analytical insights, enabling users to easily interpret and act upon the data.	*5*	UC_11_2
Req_110	iAMP-Hydro shall enable users to update and refine the pre-trained models when/if needed.	*5*	UC_11_2
Req_111	iAMP-Hydro shall develop a comprehensive strategy and framework for optimizing hydropower asset management using machine learning and IoT.	*5*	UC_11_3
Req_112	iAMP-Hydro shall incorporate historical generation data, weather patterns, river flow rates, and reservoir levels into decision support systems.	*5*	UC_11_3
Req_113	iAMP-Hydro shall employ advanced optimisation algorithms to balance objectives like maximizing energy production, minimising costs, and ensuring regulatory compliance.	*5*	UC_11_3
Req_114	iAMP-Hydro shall integrate environmental considerations and biodiversity preservation strategies into the optimisation framework.	*5*	UC_11_3

Req_115	iAMP-Hydro shall provide an advanced analytics engine supporting data preparation, machine learning, predictive analytics, and data visualisation functionalities.	*5*	UC_11_3
Req_116	iAMP-Hydro shall optimize asset management and lifecycle plans, considering factors like predicted component lifetimes, maintenance costs, and operational constraints.	*5*	UC_11_3
Req_117	iAMP-Hydro shall ensure the framework can handle new proprietary or external data sources and enhanced analytics capabilities.	*5*	UC_11_3
Req_118	iAMP-Hydro should ensure baseline algorithms are available for integration and deployment in the modular analytics framework.	*4*	UC_11_4
Req_119	iAMP-Hydro should enable seamless combination and integration of individual components into the analytics workflow by data scientists.	*4*	UC_11_4
Req_120	iAMP-Hydro should support iterative application, assessment, and refinement of analytics processes to ensure results are acceptable and significant.	*4*	UC_11_4

4 iAMP-Hydro ICT components Architecture

Following the definition of the iAMP-Hydro UCs and their associated requirements, this chapter leverages this knowledge to provide an overview of the iAMP-Hydro ICT Framework's architecture. Specifically, it delves into the conceptual architecture of the iAM Platform, outlining its core components , the IDML and IDOL and clearly defining their high-level roles within the overall iAMP-Hydro framework.

It is important to note that the architecture of the iDML builds upon the architectural design already developed in the EU's H2020 SYNERGY project (G.A. number: 872734); where the necessary modifications and extensions were elaborated to comply with the conditions of the Grant Agreement for iAMP-Hydro (G.A. number: 101122167).

Also, the official release of the iAM Platform is scheduled for M20 of the project's timeline. The intervening period will be used for further enhancements and updates to the framework, as necessary; as well as adjustments based on new findings and emerging needs as the project progresses. An illustration of the iAMP-Hydro ICT infrastructure, and the conceptual architecture of the iAM-Platform can be seen in the following figure.

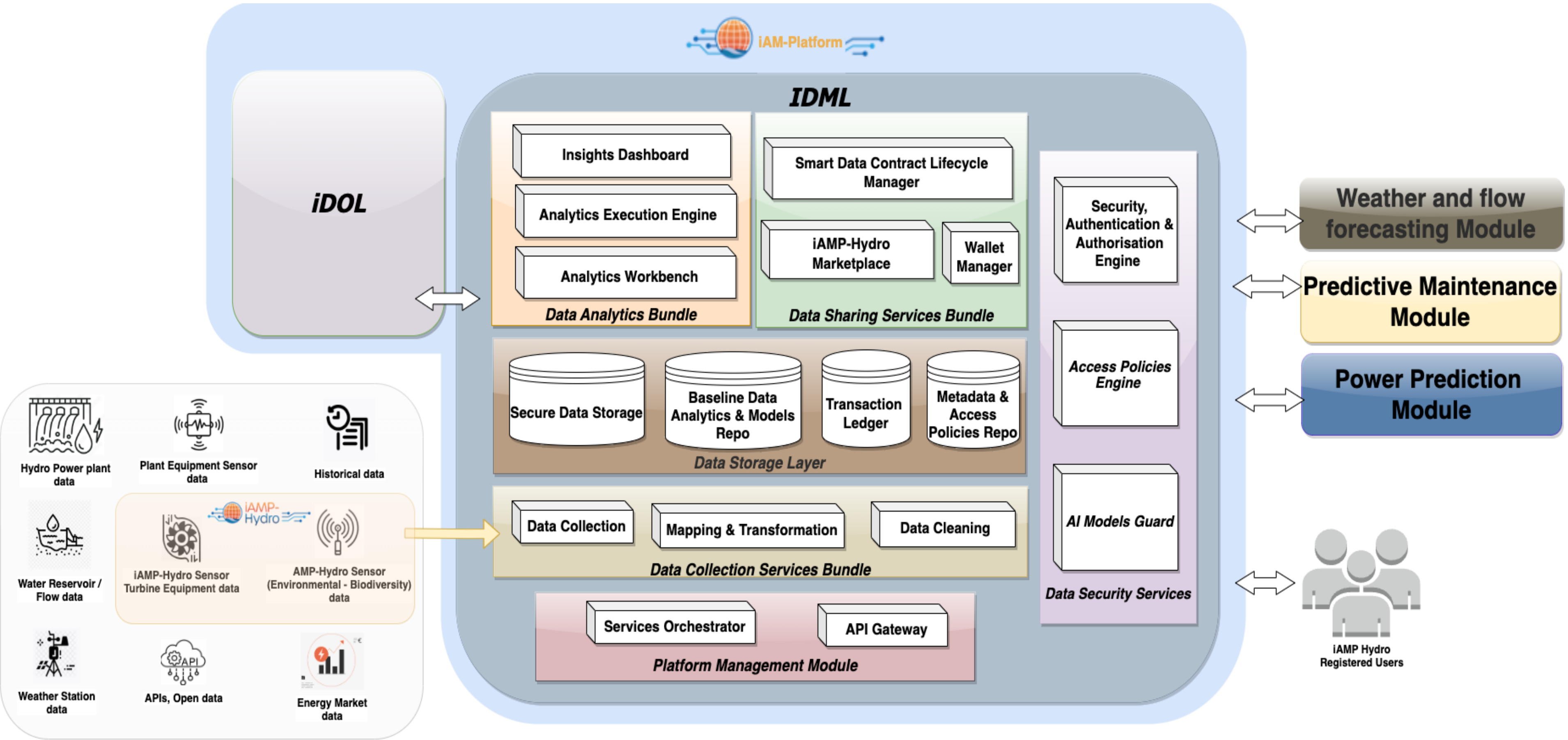

Figure 2 iAM Platform - Conceptual architecture

4.1. iDML: Services bundles, functional and technical specifications

Having provided an overview of the iAM Platform above, this section focuses on the iDML, providing a detailed examination of its' services bundles. This includes an in-depth look at their respective subcomponents, a comprehensive list of functional requirements derived from these components, and an outline of the technology stack that will be used for their development. .

Unveiling and sharing the intelligence emerging from a plethora of diverse yet disperse data assets that carry valuable energy and hydro-related information requires a novel approach. In this direction, the iDML employs four core data services bundles and a storage layer, complemented by added value services. These core bundles are namely the :

- ***Data Collection Services bundle (DCS)***: This bundle encompasses data collection (ingestion), semantic enrichment (through mapping to the project common information model, the iAMP-Hydro CIM) and transformation, along with data cleaning processes. It ensures that diverse data sources are effectively integrated into the iAMP under a common format, enabling interoperability seamless data flow and accessibility.

- ***Data Security Services bundle (DSS)***: Responsible for safeguarding and securing any data asset available through the IDML. This bundle implements robust security measures to protect data integrity and confidentiality, while also being responsible for user's access management.

- ***Data Sharing Services bundle (DShS)***: At the core of the iDML, this bundle has a dual role; It enables users to search and identify data assets of interest that are available in the iDML and allow the sharing of these data assets through the signing of smart data contract between the respective parties; while also handles and monitors the variety of smart data contracts established over a Distributed Ledger Technology (DLT) - enabled infrastructure. Overall, it enhances data assets discoverability in the iDML and ensures the secure and efficient data sharing among stakeholders.

- ***Data Analytics bundle (DA)***: This bundle allows for exploratory data analysis, design and execution of analytics workflows, and utilisation of pre-trained analytics (baseline data analytics) to generate new insights. It empowers users to derive actionable intelligence from the utilising data assets available in the iDML.

The storage and management of all artefacts available in the IDML is undertaken by the ***Data Storage Layer (DSL***), being the central secure storage component of the iDML; offering multiple storage modalities depending on the scope and types of data to be used in the iDML. It ensures that these assets are both reliably stored and retrievable by other iAMP's services and applications.

Supplementary to the above core data service bundles, the iDML incorporates also the **Platform Management Module (PMMo)**, responsible for the management and orchestration of the operations undertaken within the iDML (e.g., data collection job, data analytics workflow design, etc.), while also delivering APIs for integration with other tools and third-part services.

Overall, the iDML components' architecture is meticulously designed to facilitate the effective management of hydro-electric power, operations data, along with data from IoT devices (sensors, actuators, etc.) By leveraging these data services bundles, iDML not only enhances data integration and security, but also empowers stakeholders with valuable insights and intelligence, driving efficiency and innovation in the hydroelectric

4.1.1 Data Collection Services bundle

The Data Collection Services Bundle will feature an intuitive (UI) enabling stakeholders of the hydro-energy domain, acting as data owners/ providers, to collect, load, and prepare operational data from diverse sources to the iAM Platform; making them available for further analysis and/or to be shared with other stakeholders registered to the iAMP. The components within this bundle are divided into two categories, each serving a specific role in data handling. The following table presents the data management subcomponents included in the Data Collection Services bundle along with their mapping to the iAMP-Hydro Use Cases and associated requirements (defined in Chapter 2 and 3 respectively).

Table 21 Data Collection Services' subcomponents and mappings to iAMP-Hydro UCs &requirements

Comp. ID	Sub-component Name	Subcomponent Description	Related UC(s)	Related Reqs. (Chapter 3)
DCS	***Data Collection***	This data management component enables users to collect/upload their (structured) data from various data sources (IoT devices, sensors, actuators, external databases) and via multiple modalities, such as file uploads (.csv,.json,.xml), data collection of through Application Programming Interfaces (APIs) provided both from the iDML, or available at the data provider's side; collection of streaming data through Kafka, MQTT protocol (depending on the UC needs).	UC10_1, UC10_2	Req_020, Req_021, Req_022, Req_023, Req_024, Req_025, Req_026, Req_027, Req_028, Req_029, Req_030, Req_031
DCS	***Mapping & Transformation***	Offering an intuitive UI, enabling users to manually align (map) their diverse data formats to a common data model the IAMP-Hydro CIM, while being supported by semi-automated mappings. Users can also propose new concepts for inclusion in the CIM. This subcomponent is also responsible for undertaking all the necessary data transformations based on the user's configurations during the mapping process. Finally, it enables user to define appropriate metadata for their datasets, so as to enhance their discoverability in the iDML.	UC10_2, UC10_3	Req_032, Req_033, Req_034, Req_035, Req_036, Req_037, Req_039, Req_041, Req_042
DCS	***Data Cleaning***	This subcomponent allows users to clean their data (on demand) before uploading it to the iDML, towards enhancing data quality and ensuring completeness of the data assets, leading to easier reusability and more reliable insight extraction. It facilitates the removal or correction of incomplete, inconsistent, improperly formatted, or otherwise incorrect data. Users have access to various data cleaning functions, including value substitution, reformatting, and duplicate removal, among others.	UC10_2	Req_038, Req_039, Req_040, Req_041, Req_042

Technology stack

The different subcomponents of the Data Collection services will be built on state-of the art technologies, detailed below:

- For the *Front-end Layer*, (i.e. the users' interface), VueJS[1] and TailwindCSS[2], will be employed for a project custom front-end design.
- For the *Back-end layer* it is anticipated to utilise:

[1] https://vuejs.org
[2] https://tailwindcss.com

- Nest (NodeJS) web framework[3], which enables the development of efficient, reliable, and scalable server-side applications.
- Flask[4], a micro web framework.
- Pandas[5],that will be utilised for data handling, cleaning, and anonymisation.
- Kafka[6], a distributed stream-processing platform, which will serve as the iDML's publish-subscribe mechanism for collecting streaming data.
- RabbitMQ[7], a message broker to be utilised for exchanging feedback messages, and key management with the Platform Management Module.

4.1.2 Data Sharing Services bundle

The Data Sharing services bundle (DShS) of the IAMP enables users (i.e., data asset consumers) to search for, browse and explore data assets already available in the iAMP; and if interested proceed with its acquisition upon signing a smart data contract, supported by a distributed ledger technology (DLT). with the respective data owner. This is to ensure trusted, open and transparent data sharing practices. Overall, the Data Sharing services bundle enhances the discoverability and sharing of data assets within the iDML, enabling users to perform queries, towards identifying data assets of interest, view the resulting data assets available in the iDML along with their basic details (such as its data attributes, owner, and volume) and if interested request the acquisition of the desired data, following the data assets' predefined access policies.

The Data Sharing services bundle comprises several subcomponents, as illustrated in the iAMP's conceptual architecture (see Figure 2), and which are described in the following table along with their mapping to the project's UCs and requirements (defined in Chapter 2 and 3 respectively)

Table 22 Data Sharing n Services bundle: Subcomponents' description and mappings to iAMP-Hydro UCs and requirements

Comp. ID	Sub-component Name	Subcomponent Description	Related UC(s)	Related Reqs. (Chapter 3)
DShS	***iAMP-Hydro Marketplace***	Offers an intuitive search interface that enhances data asset discoverability and exploration. It enables users to search through the data assets available in the iAMP, browse the results (upon resolution of their access policies); explore data assets suitable for acquisition; preview their details (metadata, structure, etc.) and if satisfying their needs proceed with their acquisition by signing a smart data contract with the respective data owner. The contract is then securely stored on a distributed ledger and the component is also responsible for enforcing the respective contract terms for the data assets utilised in the iAMP. To enhance transparency and provide full traceability, all smart data contract-related information is stored in the Transaction Ledger of the Data Storage Layer.	UC_10_5, UC_10_6	Req_058, Req_059, Req_060, Req_061, Req_062, Req_063, Req_064, Req_065, Req_066

[3] https://nestjs.com
[4] https://flask.palletsprojects.com/en/2.0.x
[5] https://pandas.pydata.org
[6] https://kafka.apache.org/
[7] https://www.rabbitmq.com/

DShS	***Smart Data Contract Lifecycle Manager***	Responsible for generating the smart data contracts, and orchestrating the associated operations, including initiation of a data sharing process, creation of a machine-processable smart contract format, secure storage of the contract in a distributed ledger. Additionally, is responsible for notifying the appointed member of an organisation about the progress of data asset sharing task, where notifications are anticipated to be dispatched via email to the concerned users and through the iDML.	UC_10_6	Req_067, Req_068, Req_069, Req_070, Req_071, Req_072, Req_073
DShS	***Wallet Manager***	Responsible for managing the ledger account of IDML and interfacing with each organisation's (registered in the iDML) own Wallet. It is practically used to verify smart asset data contracts that allow an organisation to share/ obtain data.	UC_10_6	Req_073, Req_074, Req_075, Req_076

Technology stack

The fulfil its intended functionalities the Data Sharing Services will be built on state-of-the-art technologies for developing of its subcomponents. Considering that this services bundle scope is twofold:

- the *Front-end Layer* will leverage VueJS[1] and TailwindCSS[2] to enable a custom and responsive design.
- the *Back-end Layer* will utilise the NestJS (NodeJS) web framework[3] for its development serving both data asset exploration and data sharing components. PostgreSQL[15] will be employed for storing the users' query configurations, while Elasticsearch[16] will enable users to perform search queries efficiently and effectively, ensuring also robust searchability over data and metadata.
- The Blockchain Layer will utilise Ethereum's distributed platform for recording the smart data contracts, alongside existing wallet providers such as MetaMask.

4.1.3 Data Analytics Services bundle

The Data Analytics Services (DAS) bundle provides users (data providers and consumers - hydro sector stakeholders) with robust functionalities for deriving valuable insights, leveraging various analytics models and decision-making algorithms tailored to the users' requirements, on the data already integrated into the iAMP, see the results and create visualisations. Through this services bundle, data providers and consumers are empowered to design and execute sophisticated data analytics workflows using pre-existing and potentially pre-trained models and algorithms, facilitating the extraction of meaningful information. It needs to be noted that the DAS bundle of iDML will primarily support the ML/DL needs of iDOL, whereas in cases required, it encapsulates the required flexibility to support further analytics needs from the other modules involved in the iAMP-Hydro architecture (in certain cases where specific libraries are not supported by the analytics functions of these modules).

The key components of the Data Analytics Services bundle are described in the following table along with their mapping to the iAMP-Hydro UCs and associated requirements (defined in Chapter 2 and 3 respectively).

Table 23 Data Analytics Services bundle: Subcomponents' description and mapping to iAMP-Hydro UCs and requirements

Comp. ID	Sub-component Name	Subcomponent Description	Related UC(s)	Related Reqs. (Chapter 3)
DAS	***Analytics Workbench***	Offering an intuitive UI, enables users to design analytics workflows, within a structured framework. It requires data to be pre-processed into a specific format, to ensure it data can move through the various analytics functions and algorithms configured by the users. Functions supported include ready to use machine learning (ML) algorithms (such as regression, classification, clustering, and forecasting) and deep learning (DL) algorithms, aiding decision-making for iAMP-Hydro stakeholders. Users can also choose from a collection of pre-configured and trained baseline analytics workflows, namely the "Baseline Data Analytics". Users can also export their analytics workflows' results, for integration with the other iAMP-Hydro applications, thus serving various iAMP-Hydro stakeholders.	UC_11_1 UC_11_2 UC_11_3	Req_080, Req_081, Req_082, Req_083, Req_084, Req_085, Req_086, Req_087, Req_088, Req_091, Req_092, Req_093, Req_094, Req_095, Req_102, Req_115, Req_117
DAS	***Analytics Execution Engine***	Responsible for the execution of these analytics workflows to produce the relevant outcomes. In addition, this engine facilitates monitoring and safeguarding the ML/DL model transparency and efficient performance.	UC_11_1	Req_090, Req_091, Req_090, Req_093, Req_094, Req_095
DAS	***Insights Dashboard***	Enables users to visualise the results of their (executed) analytics workflows and customise how the results will be graphically displayed.	UC_11_1	Req_100, Req_102, Req_103

Technology stack

The fulfil its intended functionalities the Data Analytics Services will utilise state-of-the-art technologies for developing its subcomponents; in more detail:

- the *Front-end layer,* will be built on VueJS[1] and TailwindCSS[2], while Kibana and/or Cube.js will be used as open-source data visualisation dashboards. In respect to this service's data storage, PostgreSQL[15] will be used for the storage and retrieval of the user's configurations of the data analysis processes and the data manipulation functions; MinIO[17], for the temporary storage of the transformed data; GitLab[20] as the repository for algorithms; and MongoDB[14] for the storage of the analytics processes' results.
- the *Back-end layer* will be built on Nest (NodeJS) Web Framework[3]; Argo[8], an open source container-native workflow engine for orchestrating parallel jobs on Kubernetes; Scikit-learn[9], utilised for executing the analytics; Spark, Pandas, and Dask[10] Frameworks, to be used for data manipulation; Flask Micro Web Framework[4], will be employed for exporting the analytics results and RabbitMQ[7] for exchanging feedback messages and providing keys to the Platform Management Module.

[8] https://argo-workflows.readthedocs.io/en/latest/
[9] https://scikit-learn.org/stable/
[10] https://www.dask.org

4.1.4 Data Security Services bundle

The Data Security Services (DSS) bundle plays a pivotal role across all service bundles within the iDML, and is responsible for ensuring all platform's secure operation, while also safeguarding its resources. These resources can include data assets, AI models, the Baseline Data Analytics, the iAMP CIM, and more. The key components of the Data Security Services bundle are described in the following table along with their mapping to the iAMP-Hydro UCs and associated requirements (defined in Chapter 2 and 3 respectively).

Table 24 Data Security Services bundle: Subcomponents' description and mapping to iAMP-Hydro UCs and requirements

Comp. ID	Sub-component Name	Subcomponent Description	Related UC(s)	Related Reqs. (Chapter 3)
DSS	***Security, Authentication & Authorisation Engine***	Responsible for managing user identities and access operations for the iDML and its resources. Additionally, it provides the iDML Administrator with tools to monitor user activities, modify roles, and manage other identity-related functions. Overall, it serves as iAMP identity provider allowing for reliable registration of organisations and users, verifying an organisation and authenticating the platform's users (being organisation's members). In collaboration with the API Gateway, the Security, Authentication & Authorisation Engine verifies the API keys for external applications, while also generating tokens for the secure data exchanges between the intra-platform components.	UC_10_2, UC_10_4, UC_10_5, UC_10_6, UC_11_1	Req_032, Req_098, Req_099, Req_102 Req_098, Req_098, Req_098,
DSS	***Access Policies Engine***	This engine's role is twofold: **a)** It enables data providers (owners) during the data collection process to define their preferred access licenses to their data that will be available in the iDML, enabling full control on which stakeholders can potentially view the data assets details and potentially acquire it. **b)** Is responsible for evaluating and enforcing the access policies of each data asset. Data assets are visible to other users (through the iAMP-Hydro Marketplace) upon resolution of their respective access policies, governed by Attribute-Based Access Control (ABAC). These policies enable data providers to safeguard and share their data assets without prior knowledge of the individual data consumers within the iDML that might use their data. In the case of smart data contracts, the engine checks if a data consumer has legitimate obtained the dataset (interfacing with the Smart Data Contract Lifecycle Manager); and verifies the data assets' access policy against the data consumer's attributes. If the required attributes are missing, access	UC_10_4, UC_10_5, UC_10_6	Req_043, Req_044, Req_045, Req_046, Req_047, Req_048, Req_049, Req_050, Req_051, Req_052, Req_053, Req_054, Req_055 Req_059,

		is denied. If the consumer has obtained the data asset, the engine validates the attributes via the DLT.		
DSS	***AI Models Guard***	This subcomponent (to be delivered in the IAMP final release) is responsible for safeguarding the Machine Learning/Deep Learning models stored in the iDML against intended biases that could manipulate the training dataset and underlying algorithms. thus, lead to erroneous insights.	UC_11_1	Req_089 Req_090, Req_091, Req_089

Technology stack

The various subcomponents of the Data Security Services bundle will be built leveraging state of the art technologies, as defined below:

For the front-end layer, VueJS and TailwindCSS will be utilised as these offer custom front-end design; For the back-end layer, the Nest (NodeJS) web framework and the Casbin[11] authorisation library, and PostgreSQL[12] will be utilised as the relational database for storing the individual data assets' access policies. Moreover, for the AI Model guard, Fairness 360[13] is expected to be utilised for identifying, reporting, and mitigating discrimination and bias in ML models throughout the AI application lifecycle.

4.1.5 Data Storage Layer

The Data Storage Layer represents the central secure storage layer of the iDML and its main responsibility is the persistent storage of all the various data assets (such as datasets, analytics models, analytics results and reports, data collection job-specific data, etc.) along with their associated metadata and access policies, in a secure and reliable manner; also enabling encrypted storage of all relevant confidential data and users' credentials (e.g., tokens, API keys, usernames and passwords). Moreover, this layer enables storage of all administrative information necessary for the efficient operation and usage of the subcomponents of the iAMP. This layer has a two-fold scope; it securely stores all data assets that are collected and/or generated in the iAMP, while also makes such data assets retrievable (i.e., available) to the rest of the iAMP services and applications. As depicted in Figure 2, the Data Storage Layer is compiled by a variety of data storage and retrieval components, which are described in the following table along with their mapping to the iAMP-Hydro UCs and associated requirements (defined in Chapter 2 and 3 respectively).

Table 25 Data Storage Layer: Subcomponents description and mappings to iAMP-Hydro UCs and requirements

Comp. ID	Sub-component Name	Subcomponent Description	Related UC(s)	Related Reqs. (Chapter 3)
DSL	***Secure Data Storage***	Utilised for optimised management and storage of the data assets available in the being the outcomes of a data collection job (along with their associated metadata, enforced access policies and the data collection job's configurations), in order to be available to the other iAMP's services and tools; intermediate files, as well as users'	UC_10_1 UC_10_2 UC_10_4 UC_10_5	Req_020, Req_028, Req_032, Req_048, Req_051, Req_056

[11] https://casbin.org

[12] https://www.postgresql.org/

[13] https://aif360.res.ibm.com

		credentials (usernames, passwords), tokens , APIs' keys, etc.		
DSL	***Baseline Data Analytics & Models Repo***	Utilised for optimised management and storage of the results of the analytics job (such as analytics models, results along with their reports) and the Baseline Data Analytics. This is a collection of pre-configured and trained baseline analytics workflows, which cover areas like predictive maintenance, biodiversity, weather, and flow forecasting; with the objective to streamline the overall process, minimising the time required to design, set up and execute such analytics workflows. This storage component is also responsible for storing the various versions of the CIM (output of T9.1), enabling thorough communication with the Mapping & Transformation for the mapping of the different information entities to the iAMP CIM concepts (see section 4.1.1).	UC_11_1 UC_11_2 UC_11_3	Req_080, Req_081, Req_082, Req_083, Req_084, Req_085, Req_086, Req_087, Req_088, Req_089, Req_088, Req_097, Req_098, Req_099, Req_102, Req_103, Req_104, Req_105, Req_106, Req_107, Req_108, Req_110, Req_118
DSL	***Transaction Ledger***	Responsible for storing the various data asset contract-related information, while securing the privacy of the involved stakeholders and providing full traceability.	UC_10_6	Req_066, Req_067, Req_073
DSL	***Metadata & Access Policies Repo***	Responsible for the storage of the data assets (available in the iDML) metadata as well for their respective access policies and define by their respective data owners.	UC_10_3 UC_10_4 UC_10_5 UC_10_6	Req_041, Req_047, Req_049, Req_050, Req_052, Req_053, Req_054, Req_055, Req_059, Req_060

Technology stack

The Data Storage Layer will be built on state-of the art indexing and storage technologies, detailed below:

- MongoDB[14] as the NoSQL database for the storage of the various data assets uploaded in the iDML,
- PostgreSQL[15], as the iDML's relational database,
- Elasticsearch[16], will be utilised as the search optimisation and indexing engine,
- MinIO[17], (object repository) as the temporarily storage data lake,

[14] https://www.mongodb.com/
[15] https://www.postgresql.org/
[16] https://www.elastic.co/
[17] https://min.io/

- Vault[18] will be utilised for storing sensitive data, (i.e., usernames, passwords, tokens, APIs Keys, etc.\),
- the transaction ledger technology will be based likely on Ethereum[19],
- Gitlab[20], which will be used for the storage of the baseline algorithms.

4.1.6 Platform Management Module

The Platform Management Module (PMMo) integrates essential security and system management components necessary for the iDML's efficient operation. The key sub-components of the Platform Management Module are described in the following table along with their mapping to the iAMP-Hydro UCs and associated requirements (defined in Chapter 2 and 3 respectively).

Table 26 Platform Management Module: Subcomponents description and mappings to iAMP-Hydro UCs and requirements

Comp. ID	Sub-component Name	Subcomponent Description	Related UC(s)	Related Reqs. (Chapter 3)
PMMo	***Services Orchestrator***	Ensures seamless operation of the iDML, specifically managing the data ingestion and analysis jobs. It allocates the appropriate computing and storage resources based on data volume and services' needs for each task execution. In addition, is in charge of delivering timely alerts to users of the iDML, keeping them informed about various activities relevant to their interests, e.g. updates on the status of ongoing data collection tasks and/or data analysis tasks.	UC10_1, UC10_2, UC10_3, UC10_4, UC10_5, UC10_6, UC11_1	Req_079
PMMo	***API Gateway***	Functions as the primary access point for all applications and stakeholders (including the iAMP-Hydro tools/modules), who are seeking data assets or analytics results within the iDML. It allows authorised stakeholders to retrieve data through iDML's APIs using specified configurations and parameters. The API Gateway sends the requests to the relevant services, components such as the Data Storage layer and Access Rights Engine, to ensure data security and proper access control.	UC10_1 UC10_6	Req_024, Req_025, Req_026, Req_027, Req_077, Req_078

Technology stack

The Platform Management Module's subcomponents will fulfil the intended functionalities deriving from the specified requirements by utilising various advanced technologies for their implementation. The technologies to be leveraged include:

- Nest (NodeJS) Web Framework[3] enabling the creating efficient, reliable, and scalable server-side applications
- Keycloak[21] which is an open-source identity provider for user access management.

[18] https://www.vaultproject.io/
[19] https://ethereum.org/en/
[20] https://about.gitlab.com/
[21] https://www.keycloak.org

- Kubernetes[22]; a portable platform for managing containerised workloads and services across different Kubernetes clusters, along with Docker for containerising the various IDML's services.

4.1.7 iDML - Non-Functional requirements

In addition to the description of the various services and subcomponent of the iDML, this section presents a non-exhausting list of non-functional requirements for the iDML.

Table 27 iDML - Non-Functional requirements

Req. _ID	Description
NFREQ_01	The iDML shall support user registration and login using credentials.
NFREQ_02	The iDML shall deliver to its users an intuitive and user-friendly user interface(UI).
NFREQ_03	The iDML shall provide responsive design to ensure usability across various devices and screen sizes.
NFREQ_04	The iDML shall be accessible via a web browser.
NFREQ__0	The iDML's UI shall be available in the English Language as a minimum.
NFREQ_06	The iDML may be provided in the official language of iAMP-Hydro demo sites.
NFREQ_07	The iDML shall maintain data integrity and accuracy throughout data processing and storage.
NFREQ_08	The iDML shall ensure data security and privacy in compliance with industry standards.
NFREQ__0	The iDML shall be interoperable, allowing its services to interface seamlessly with other iAMP Hydro module, systems, etc.
NFREQ_10	The iDML shall provide comprehensive logging and auditing features for critical user and system activities.
NFREQ_11	The iDML shall support scalable data storage to handle large volumes of hydro-energy, sensor data
NFREQ_12	The iDML shall offer a robust APIs for data integration and third-party applications.
NFREQ_13	The iDML shall provide automated backup and recovery mechanisms.
NFREQ _14	The iDML include disaster recovery capabilities to ensure data and service continuity in case of major failures.
NFREQ _15	The iDML shall ensure high availability and minimal downtime
NFREQ _16	The iDML shall support role-based customisation of user interfaces to reflect different user roles and responsibilities.
NFREQ _17	The iDML shall have a modular architecture to facilitate easy updates and integration of new features
NFREQ _18	The iDML shall offer scalable compute resources to dynamically adjust to varying workloads and user demands.

4.1.8 iDML User Workflows

Having described the key components and services of the iDML, this section entails the three key user workflows anticipated over the iDML and which consist of the:

1. ***Data Collection workflow***, which includes the process that data providers shall follow in order to upload their datasets to the iDML and upon processing them, make them available onto the iDML to be used for their own organisation's purposes and/or for sharing them.
2. ***Data Analytics Design-Execution workflow***, *which* outlines the process through which users of the iDML can leverage the data assets they own (and which are already available in the iDML), to design and execute analytic workflows for extracting valuable insights and intelligence to support their organisation's decision-making.

[22] https://kubernetes.io

3. ***Data Asset Querying and Sharing workflow***, which entails the process by which users (data consumers) can search for and access/acquire data assets available within the iDML, through the signing of smart data contracts with the respective data owners (provides), upon resolution of the respective data asset's access policies.

These core workflows are described in more detail in the following sections as follows:

4.1.8.1 Data Collection workflow

As shown in the figure below, depicting the overall data collection process and involved subcomponents/layers; the process initiates with a Data Provider/Owner registering and signing-in to the iDML with the intention to upload his/her data and make them available on the iDML for private use within his/her organisation, or for sharing with other registered organisations.

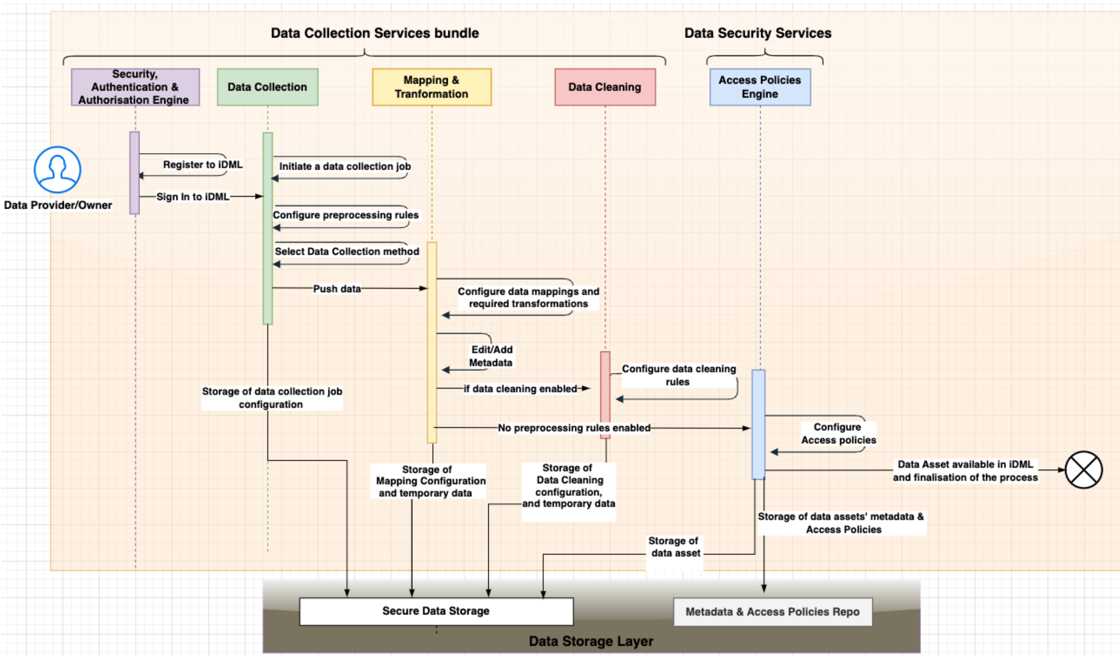

Figure 3 Data Collection Workflow

Upon successfully signing-in to the iDML, the users (acting as data providers) can upload their data via different modalities, and this is done by creating a new Data Collection Job (DCJ) via the Data Collection component. As a first step, the user configures his/her preferred preprocessing steps, such as cleaning, of his/her data, prior to being to being uploaded into the iDML. Following the user can selects his/her preferred way of upload his/her data, via the different modalities to be offered by the iDML. These can include direct file upload (e.g., .json, .xml, etc.), upload data through the data provider's own APIs, or using the iDML 's APIs.

Once the Data Collection Job is properly set up, the user shall upload a sample of the data, as well as the whole file. If the data collection process completes successfully, the data are stored temporarily; however, if any issues are encountered, any error details are recorded, and the respective DCJ is marked as failed. Appropriate feedback messages are provided to the user, supporting him/her in identifying the issue(s) and take corrective action.

Upon successful completion of a DJB, the Services Orchestrator activates the execution of the Mapping and Transformation service (see section 4.1.1) in the iDML, providing an interface to the user to manually map the attributes (e.g. columns) of their datasets to the attributes (concept and fields) of the iAMP-Hydro CIM. Semi-automatic mappings are also provided to the users to assist them in mapping (all) the attributes of their dataset that they want to upload in the iDML. In this step user shall also define the preferred transformations to be applied on their data (e.g., change of measurement units, etc.). Upon successful completion of the data mapping process, the transformed data are temporarily stored, else the related error information is collected and the specific step of the respective DCJ is marked as failed.

If the cleaning step is enabled in the Data Collection Job configuration, the Services Orchestrator triggers the execution of the Data Cleaning component. Here, the user can define his/ger preferred cleaning rules to be applied to their data set prior to being uploaded to the iDML. As previously, If the data cleaning process is completed successfully, the cleansed data are temporarily stored, else the related error information is collected and the specific step of the respective DJB is marked as failed.

Finally, through the Access Policies Engine, the user is enabled to define his/her preferred access policies for his/her dataset (that will be enforced through its lifecycle), also proving appropriate metadata. This workflow concludes with the storage of the data providers data asset in the secure data storage of the iDML.

4.1.8.2 Data Analytics Design-Execution workflow

The Data Analytics Design-Execution workflow outlines the process through which users can use the data assets they own, and which are available in the iDML, towards designing and executing analytic workflows for extracting valuable insights and intelligence to support their organisation's decision-making.

Considering that a user has already signed-in to the iDML, through the interface of the Analytics Workbench (available in the Data Analytics Bundle, see 4.1.3) a user can create a new data analytics workflow. This entail, selecting the data assets he/she wishes to analyse (from those owned by his/her organisation) and proceeding with the configuration of the desired parameters. For example, specifying the types of analyses to be performed (ML, statistical, etc), selecting the algorithms from a predefined list of ready-to-use baseline analytic algorithms, etc... Once the data analytics workflow is configured, the user can start its execution immediately or schedule it for automatic execution at predetermined times. This includes tasks like data processing, model training, or creating visualizations based on the user's settings. Upon successful execution the results are made available through the Insights Dashboard (available in the Data Analytics Bundle, see 4.1.3), offering an interface for presenting the results in formats predefined during the design phase, such as reports, dashboards, or data visualisations. The user can review the insights and findings derived from the analysis. Depending on the results, they may choose to adjust the analytic workflow parameters or re-execute the analysis to refine the insights or explore other aspects of the data.

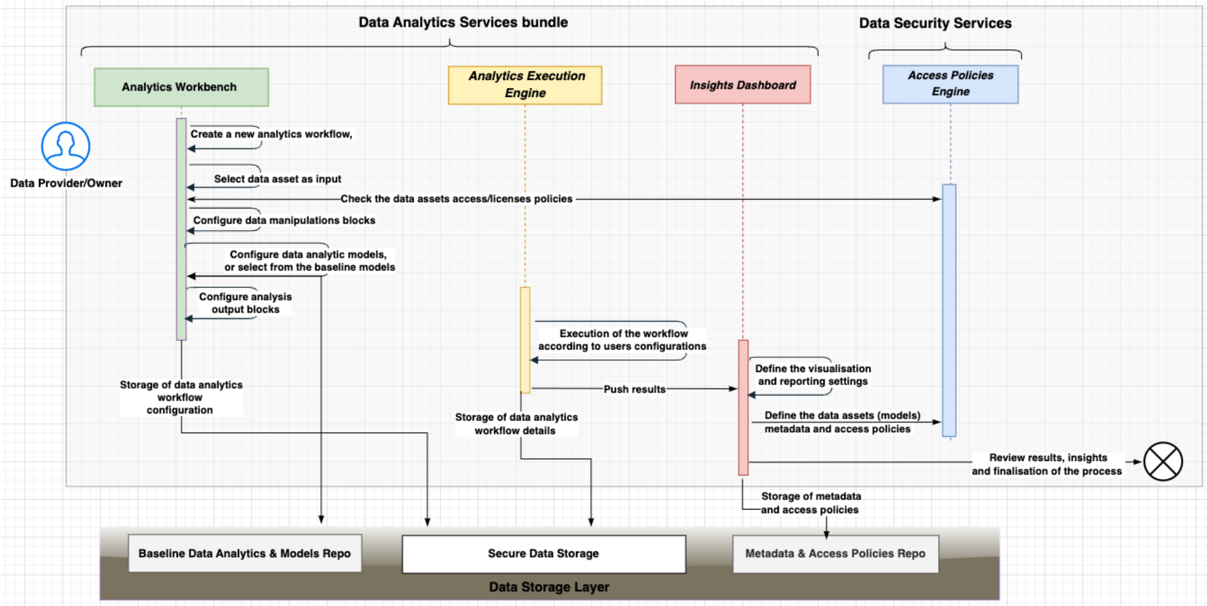

Figure 4 Data Analytics Design-execution Workflow

4.1.8.3 Data Asset Querying and Sharing workflow

The Data Asset Querying and Sharing workflow details the process by which users can search for and access/acquire data assets available within the iDML, through the signing of smart data contracts with the respective data owners.

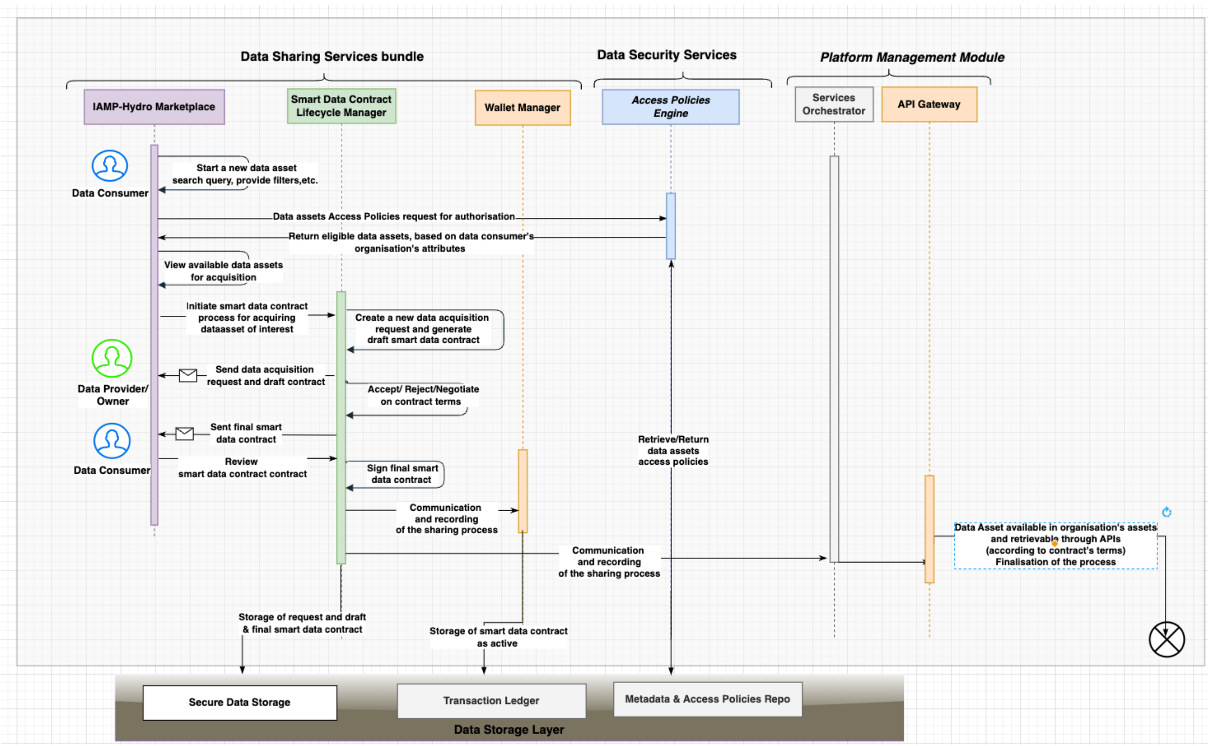

Figure 5 Data Asset Querying and Sharing Workflow

As depicted in the figure above, which outlines the overall process and involved subcomponents/layers, this workflow begins with a user wishing to explore data assets available on the iDML and which can be of interest to his/her organisation's operations. This can be done by accessing the iAMP-Hydro Marketplace (available in the Data Sharing Bundle, see 4.1.2). Here, they can browse through the results or apply filters to find specific data assets based on criteria such as keywords, data types, metadata, or tags. The user configures the search parameters by specifying the criteria and filters to narrow down the results. This includes setting up search keywords, selecting data categories, or defining other relevant attributes. The user executes the query, and the iDML performs a search across the available data assets in the iDML, displaying in the iAMP-Hydro Marketplace the results in a user-friendly format; upon resolution of their access policies by the Access Policies Engine. The search query results that do not violate any access policy are displayed to the data asset consumer in the iAMP-Hydro Marketplace. From the search results, the user selects the data assets they are interested in. They can preview details of the selected assets, such as metadata, descriptions, and sample data. If the user has the necessary permissions, they can initiate a data sharing process by initiating of Smart Data Contract towards eventually accessing and/or downloading the data assets for their own use.

As next step, the respective Data Provider/Owner is notified and by accessing the iAMP-Hydro Marketplace he/she can review the details of the request for quotation, as well as of the organisation that made the request, and decide whether to accepted it or not. If the request is rejected, the Data Consumer is notified accordingly. If the request is accepted, the Data Provider prepares a draft smart data asset contract with the help of the Smart Data Contract Lifecycle Manager that is stored in the blockchain (Transaction ledger) in the Data Storage Services.

The Data Consumer is notified for the draft contract and accesses the Smart Data Contract Lifecycle Manager (via the iAMP-Hydro Marketplace interface) to review the contract's details; if the terms of the contract are acceptable, he/she can shall proceed with signing the contract; if he/she does not agree with the terms he/she can reject it; if the terms are partially acceptable, the data Consumer can enter into negotiations with the Data Provider/Owner, which can become an iteratively with counter-offers from each party, all documented as different versions of the smart data contract and which are all stored in the Transaction ledger (blockchain). Once there is agreement on the smart data contracts' terms, both parties shall sign the final version of the smart data contract so that it can become active, and the sharing of the respective data assets is allowed.

During this process, the iDML also processes (through the Services Orchestrator) the sharing requests and updates the access permissions for the specified data assets accordingly. Once the data assets are shared, the recipients are notified of the new access permissions and can view or utilise the shared assets based on the smart data contract terms. Once the data asset consumers have an active data asset contract, they are also able to define how they wish to retrieve the data, e.g. as a file, or through an AP, offered by the API Gateway. For the latter case, the user needs to configure the retrieval settings.

Overall, this workflow facilitates efficient querying and sharing of data assets within the iDML, enhancing collaboration and data accessibility, while maintaining appropriate access controls and user permissions.

iDOL: functional and technical specifications

The following section focuses on iDOL (Integrated Decision Optimization Layer) and offers an examination of its service bundles. The iDOL is designed to integrate outputs from various work packages (WPs) and external data inputs, such as energy prices, to form a robust optimization layer. This layer enhances operational efficiency and decision-making processes within the hydroelectric sector by leveraging diverse data assets and advanced algorithms. In this direction, the iDOL employs several core optimization bundles, each with specific functions and purposes:

- **Data Collection and Aggregation Bundle (DCAB):** The Data Collection and Aggregation Bundle (DCAB) is essential for continuously gathering and integrating information from a variety of data sources. This bundle ensures that all relevant data is collected, aggregated, and made accessible for further analysis and optimization within the iDOL framework.
- **Integration Bundle (IB)**: This bundle is designed to integrate and utilise data from various digital solutions, including predictive maintenance, ecological status monitoring, and weather and flow forecasting. By compiling outputs from different work packages (WPs), IB provides a comprehensive approach to hydroelectric operations.
- **External Data Integration Bundle (EDIB):** This bundle ensures seamless incorporation of external data sources, enriching the optimization process with relevant information. EDIB integrates data from external sources such as energy prices, regulatory updates, and market trends.
- **Real-Time Data Aggregation and Analysis Bundle (RTDAAB)**: This bundle aggregates real-time data from internal and external sources for comprehensive analysis. RTDAAB provides a holistic view of hydroelectric operations, enhancing decision-making through real-time data dashboards, analytical reports, and actionable insights. It ensures data integration from various WPs and external datasets, utilizing statistical and machine learning techniques.
- **Statistical and Machine Learning Core Bundle (SMLCB):** The Statistical and Machine Learning Core Bundle (SMLCB) plays a crucial role in optimizing hydroelectric power operations by leveraging advanced data analysis techniques. This bundle utilises a combination of methodologies and algorithms to enhance various aspects of operational efficiency, from predictive maintenance to energy production optimization.

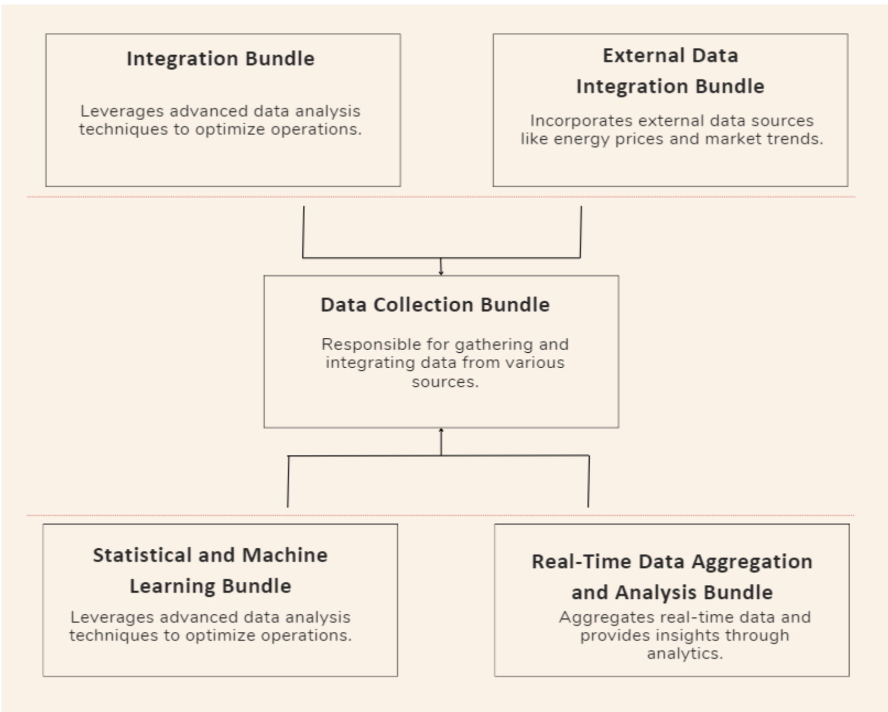

Figure 6 IDOL internal Architecture

4.1.9 Data Collection and Aggregation Bundle

The Data Collection and Aggregation Bundle is essential for continuously gathering and integrating information from a variety of data sources. This bundle ensures that all relevant data is collected, aggregated, and made accessible for further analysis and optimization within the iDOL framework. The components within this bundle focus on data management, each with a specific role in data handling.

Technology stack:

The Data Collection and Aggregation bundle will utilise Pandas[5] for in-memory data processing.

4.1.10 Integration Bundle

The Integration Bundle (IB) enhances hydroelectric power operations by leveraging data from predictive maintenance, ecological status monitoring, and weather and flow forecasting. It compiles outputs from different work packages (WPs) to optimize operations comprehensively. Predictive maintenance anticipates and prevents equipment failures through predictive models, minimizing downtime and extending asset lifespan. Ecological status monitoring ensures compliance with sustainability goals by collecting and analysing environmental sensor data. Weather and flow forecasting integrates weather forecasts and hydrological models to optimise energy production. By integrating these functions, the IB provides a data-driven approach to enhancing predictive maintenance, ecological monitoring, and operational efficiency, promoting sustainability and reliability in hydroelectric power operations.

Table 28 IB subcomponents description and mappings to iAMP-Hydro UCs and requirements

Comp. ID	Sub-component Name	Subcomponent Description	Related UC(s)	Related Reqs. (Chapter 3)
IB_1	***Data Aggregator***	Aggregates data from various sources such as sensors, historical maintenance records, and operational logs for analysis.	UC_2_1, UC_3_1, UC_4_1, UC_6_1	Req_002, Req_010, Req_017, Req_020

Technology stack

The subcomponent of the iDOL will utilise Scikit-learn[9] for predictive models, keras[23] or tensorflow[24] for deep learning, sktime[25] for time series models.

4.1.11 External Data Integration Bundle

The External Data Integration Bundle (EDIB) ensures seamless incorporation of external data sources, enriching the optimization process with relevant information. This bundle integrates data from external sources such as energy prices, and market trends, providing a comprehensive view that enhances decision-making and operational efficiency within the iDOL framework. The following table presents the external data integration subcomponents included in the External Data Integration Bundle along with their mapping to the iDOL's UCs and associated requirements:

Table 30 EDIB subcomponents description and mappings to iAMP-Hydro UCs and requirements

Comp. ID	Sub-component Name	Subcomponent Description	Related UC(s)	Related Reqs. (Chapter 3)

[23] https://keras.io

[24] https://www.tensorflow.org

[25] https://www.sktime.net/en/stable/

EDIB_1	***Market Data Collector***	Gathers data from external sources such as energy prices, renewable energy generation historical data, and market trends to inform optimization models.	UC_11_1	Req_017 Req_020
EDIB_2	***External Data Aggregator***	Aggregates data from various external sources, ensuring that it is normalized and ready for integration into the iDOL. This includes data from APIs, such as weather forecasts.	UC_11_1	Req_017 Req_020
EDIB_3	***Market Trend Analyzer***	Analyses market trends and develops time series forecasts using historical data collected by EDIB_1 and EDIB_2. This includes forecasting market prices and renewable energy generation for the next week.	UC_11_1	Req_017 Req_020

Technology stack

The subcomponents of the iDOL will be utilising the following libraries depending on the intended operation:

- API Integration : Requests for API data collection, pandas for handling tabular data.
- Market Data Handling: Use pandas-datareader[26] to fetch energy prices and market trends.
- Data Aggregation: Python's pandas[5], numpy[36], or dask[27] for aggregating large datasets from multiple sources.

4.1.12 Real-Time Data Aggregation and Analysis Bundle

The Real-Time Data Aggregation and Analysis Bundle (RTDAAB) aggregates real-time data from internal and external sources for comprehensive analysis. RTDAAB provides a holistic view of hydroelectric operations, enhancing decision-making through real-time data dashboards, analytical reports, and actionable insights. It ensures data integration from various WPs and external datasets, utilising statistical and machine learning techniques. The following table presents the real-time data aggregation and analysis subcomponents included in the RTDAAB along with their mapping to the iDOL's UCs and associated requirements:

Table 31 RTDAAB subcomponents description and mappings to iAMP-Hydro UCs and requirements

Comp. ID	**Sub-component Name**	**Subcomponent Description**	**Related UC(s)**	**Related Reqs. (Chapter 3)**
RTDAAB_1	***Real-Time Data Collector***	Aggregates real-time data from various internal (e.g., sensors, SCADA systems) and external (e.g., weather data) sources.	UC_2_1, UC_3_1, UC_4_1, UC_6_1	Req_002 Req_010 Req_017
RTDAAB_2	***Data Integration Engine***	Ensures seamless integration of data from multiple WPs and external datasets, maintaining consistency and accuracy.	UC_2_1, UC_3_1, UC_4_1, UC_6_1, UC_11_1	Req_002 Req_010 Req_017

[26] https://pandas-datareader.readthedocs.io/en/latest/#
[27] https://www.dask.org

RTDAAB_3	***Reporting Module***	Generates analytical reports and actionable insights based on real-time data analysis, supporting informed decision-making.	UC_11_2, UC_11_3, UC_11_4	Req_096

Technology stack

Depending on the scope, the RTDAAB will utilise the following libraries:

- Data Analysis: pandas[5] and numpy[36] for data manipulation, matplotlib[34], plotly[28] for real-time visualisation.
- Machine Learning: scikit-learn[9] for statistical models, tensorflow[24]/keras[23] for deep learning.

4.1.13 Statistical and Machine Learning Bundle

The Statistical and Machine Learning Bundle (SMLB) plays a crucial role in optimizing hydroelectric power operations by leveraging advanced data analysis techniques. SMLB combines methodologies and algorithms to enhance various aspects of operational efficiency, from predictive maintenance to energy production optimization. The following table presents the core statistical and ML subcomponents included in the SMLB mapped to the project's UCs and associated requirements.

Table 32 SMLB subcomponents description and mappings to iAMP-Hydro UCs and requirements

Comp. ID	Sub-component Name	Subcomponent Description	Related UC(s)	Related Reqs. (Chapter 3)
SMLB_1	***Energy Production Optimizer***	Applies statistical models to optimize energy production based on real-time and historical data.	UC_11_2, UC_11_3	Req_099 Req_100 Req_103
SMLB_2	***Performance Analysis Tool***	Analyses equipment and system performance data to identify inefficiencies and recommend improvements.	UC_11_2, UC_11_3	Req_092 Req_093 Req_102 Req_103
SMLB_3	***Data Preprocessing Module***	Prepares data for analysis by cleaning, normalizing, and transforming it into suitable formats for machine learning models.	UC_11_2, UC_11_3	Req_036 Req_037 Req_038 Req_039 Req_040

Technology stack

The Statistical and Machine Learning Bundle will utilise the following libraries:

- Data Preprocessing: pandas[5] and numpy[36] for cleaning and transforming data, scikit-learn[9] for data normalisation and scaling.
- Statistical Models: statsmodels[29] for regression models, scikit-learn[9] for classification/clustering.
- Machine Learning: scikit-learn[9] for traditional ML models, tensorflow[24] or pytorch[30] for more complex neural networks.
- Optimization Tools: scipy.optimize[31] for mathematical optimization.

[28] https://plotly.com

[29] https://www.statsmodels.org/stable/index.html

[30] https://pytorch.org

[31] https://docs.scipy.org/doc/scipy/tutorial/optimize.html

4.2 Predictive Maintenance Module: Functional and technical specifications

Overview of Predictive Maintenance Module

The Predictive Maintenance Module (PMM) of the hydropower turbines within the *iAM Platform* is designed to optimize the operational efficiency and longevity of hydropower facilities. This module utilises advanced data-driven techniques and state-of-the-art sensor technology to monitor, analyse, and predict the functioning state and performance of turbine components. By leveraging real-time data and ML algorithms, the module enables timely maintenance interventions, reducing downtime and enhancing the overall reliability of the ageing hydropower system in Europe.

Functional Specifications

The Predictive Maintenance Module serves several critical functions that contribute to the optimal operation of hydropower turbines:

- **PdM1:** Real-time monitoring bundle
  - Continuous tracking of turbine parameters such as vibration, acoustic emission, pressure and velocity fluctuations, and internal flow characteristics.
  - Utilisation of sensors to gather comprehensive data, providing a holistic view of the turbine's health.
- **PdM2:** Data analysis bundle
  - Application of machine learning algorithms and statistical models to analyse historical and real-time data.
  - Identification of patterns and anomalies that may indicate incipient faults or potential failures or inefficiencies.
- **PdM3:** Predictive analytics bundle
  - Detecting faults and identifying failure points using predictive algorithms.
  - Estimation of the Remaining Useful Life (RUL) of turbine components which could facilitate to prioritize maintenance tasks.
- **PdM4:** Reporting and visualization bundle
  - Generation of detailed reports and dashboards that present turbine health metrics, and performance trends.
  - Visualization tools for easy interpretation of complex data, aiding decision-makers in strategic planning.
  - User-friendly interface providing real-time alerts and notifications to operators.

4.2.1 Predictive Maintenance Module: Subcomponents description

The key components of the Predictive Maintenance Module are illustrated in the following figure and described in the following tables along with their mapping to the iAMP-Hydro UCs and associated requirements (defined in Chapter 2 and 3 respectively).

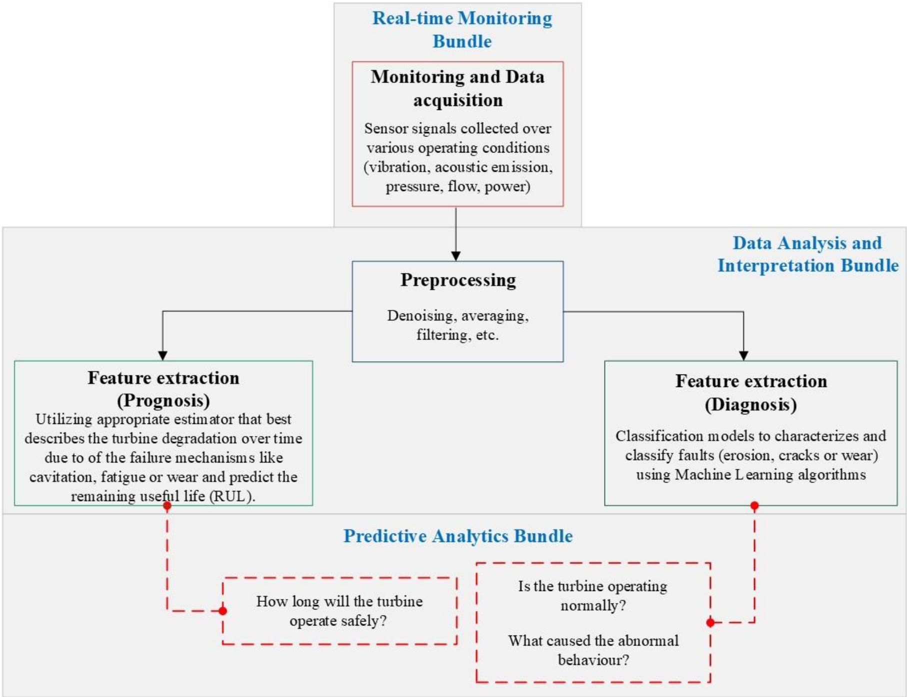

Figure 7 Predictive Maintenance Module internal architecture

Table 29 Predictive Maintenance Module: Subcomponents description and mappings to iAMP-Hydro UCs and requirements

Comp. ID	Sub-component Name	Subcomponent Description	Related UC (s)	Related Reqs. (Chapter 3)
PMM	PdM1	Real-time Monitoring bundle *(see detailed description in functional specifications section 4.3 above)*	UC_2_1	Req_001, Req_002, Req_003, Req_ 004
PMM	PdM2	Data Analysis and Interpretation bundle *(see detailed description in functional specifications section 4.3 above)*	UC_3_1	Req_002, Req_005, Req_006, Req_007, Req_008
PMM	PdM3	Predictive Analytics bundle *(see detailed description in functional specifications section 4.3 above))*	UC_3_1	Req_006, Req_007, Req_008
PMM	PdM4	Reporting and Visualization bundle *(see detailed description in functional specifications section 4.3 above)*	UC_3_1	Req_007, Req_008

4.2.2 Data analysis bundle

The Data analysis bundle is a core component of the PMM in the iAM Platform for hydropower turbines. It uses advanced computer algorithms to transform raw data from condition monitoring sensors installed on the turbines to get actionable insights. The analytics algorithm identifies deviation patterns of vibration, acoustic emission and pressure data from the healthy state facilitating prior identification of potential problems that could lead to failure of turbine components. By spotting these issues early, the system allows for proactive maintenance. This reduces downtime and improves the overall reliability and performance of the hydropower system. The analysis also helps plan maintenance more efficiently, extending the life of the aging hydropower equipment and enhancing its efficiency.

In simpler terms, this data analysis bundle gives hydropower operators the information they need to stay on top of potential problems with their turbines. This helps them maintain the turbines better and keep the hydropower plants running smoothly for longer.

The key components of the Data analysis bundle are described in the following table along with their mapping to the iAMP-Hydro UCs and associated requirements (defined in Chapter 2 and 3 respectively).

Table 30 Data Analysis bundle: Subcomponents description and mappings to iAMP-Hydro UCs and requirements

Comp. ID	Sub-component Name	Subcomponent Description	Related UC(s)	Related Reqs. (Chapter 3)
DA_1	***Application of Machine Learning Algorithms***	Utilises advanced machine learning algorithms to analyse historical and real-time data from condition monitoring sensors.	UC_3_1	Req_002, Req_003, Req_004, Req_005, Req_006
DA_2	***Statistical Models***	Employs statistical models to interpret data trends and variations over time.	UC_3_1	Req_005
DA_3	***Pattern Identification***	Identifies patterns in the data that may indicate normal or optimal operation states.	UC_3_1	Req_006, Req_007
DA_4	***Anomaly Detection***	Detects anomalies that could signal incipient faults, potential failures, or inefficiencies in turbine components.	UC_3_1	Req_006
DA_5	***Historical Data Analysis***	Examines historical data to understand long-term trends and past performance of turbine components.	UC_11_1	Req_001, Req_002

4.3 Weather & Flow Forecasting Module: functional and technical specifications

The Weather and Flow Forecasting Module (WFFM) is designed to predict the available water within different validation sites. This module forecasts the available water (the river flow / the inflow) in the dam section. The forecasting models utilise historical flow/volume data alongside climate variables such as precipitation and temperature. It shall be noted that these models will be developed and trained within WP6 and WP7 and be described in detail in the associated WP6 deliverable, D6.2 (due in M18).

A ".pkl" file will be generated for ML models, and equivalent files for physical models will also be deployed from these work packages. The models will require input from the rest of the platform for their execution. This input includes a validation site identifier, which facilitates a database connection. A set of variables specified in a configuration file will execute this database connection and run the model.

The Configuration File can also retrieve additional information needed from the end user, addressing needs that may arise during the development of the individual models.

Results from the weather and flow forecasting module can be provided in various graphical formats, such as PNG, JPEG, or SVG. Additionally, a .csv file or a data frame will be generated to facilitate the execution of the energy prediction phase.

The following figure illustrates a schematic representation of the general architecture for the Weather and Flow Forecasting module.

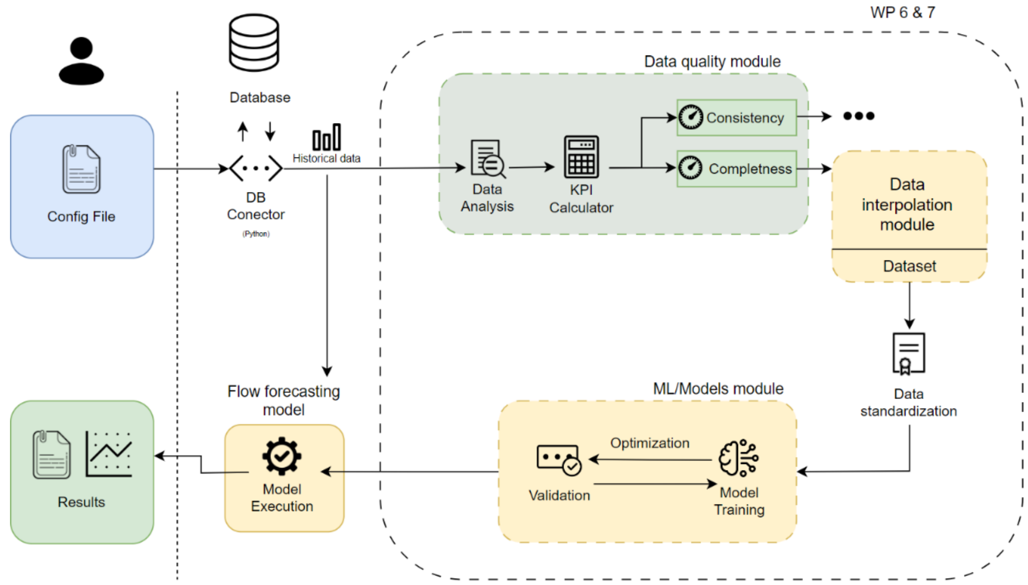

Figure 8 Weather and flow forecasting module general architecture

4.3.1 Weather & Flow Forecasting Module: Subcomponents description

The key components of the Weather & Flow Forecasting Module are described in the following table along with their mapping to the iAMP-Hydro UCs and associated requirements (defined in Chapter 2 and 3 respectively).

Table 31 Weather & Flow Forecasting Module*: Subcomponents description and mappings to iAMP-Hydro UCs and requirements*

Comp. ID	Sub-component Name	Subcomponent Description	Related UC(s)	Related Reqs. (Chapter 3)
WFFM	Config file	Component designed to manage the operations and inputs required for the overall Weather and Flow forecasting module as well as the Power Prediction module.	UC_6_1	-
WFFM	Database connector	Handles the connection with the IDML to retrieve historical data, including flow/volume data, climate variables (precipitation, temperature), historical energy production data, water level data, and turbined water data. This connection is essential for accessing the input data necessary for the model execution.	UC_6_1, UC_6_3	Req_017, Req_018
WFFM	Data Quality Module	This module is designed to prepare the dataset for the training of the models. The process involves selecting the appropriate time interval from the historical data, removing duplicate entries, and eliminating outliers. Additionally, the module identifies and selects the relevant variables and formats the data according to the requirements for model training.	UC_6_1 UC_6_3	Req_015
WFFM	Data interpolation module	The data interpolation module is intended to fill in missing data between samples for the relevant variables.	UC_6_1 UC_6_3	Req_015
WFFM	ML module	The ML module is responsible for training models using the pre-processed data and validating their performance. This module also involves selecting the best features for optimal model performance. The outcome of this process is a trained model ready for deployment	UC_6_1 UC_6_3	Req_012, Req_013, Req_014, Req_015
WFFM	Flow forecasting model	The model components, developed and trained within WP6 and WP7, are encapsulated in .pkl files or other equivalent formats. These models use data retrieved in the database connect as input data to perform predictions.	UC_6_1 UC_6_3	Req_012, Req_013, Req_014, Req_015

Technology stack

The development and execution of the Weather and Flow Forecasting Module, rely on a versatile technology stack that will vary according to the project's validation sites. The stack for the CUERVA Validation Sites includes various Python libraries and tools that facilitate data processing, model development, training, and visualisation:

- Data Processing and Management:
  - Pandas[5]: Used for data manipulation and analysis, providing data structures for handling datasets.
- Machine Learning and Model Development:
  - Scikit-learn[9]: Library for machine learning that includes tools for data preprocessing, model selection, and evaluation, implementing algorithms such as Random Forests, Support Vector Machines (SVR), Linear Regression, Decision Trees, and more.

- XGBoost[32]: A gradient boosting library designed for high performance and efficiency.

- Model Persistence:
  - Joblib[33]: A tool for serializing Python objects, used to save and load trained models.
- Data Visualisation:
  - Matplotlib[34]: A plotting library used for visualizations in Python.
  - Seaborn[35]: A data visualization library based on Matplotlib.
- Model Evaluation:
  - Scikit-learn Metrics[9]: Tools for evaluating model performance, including metrics such as mean squared error and R^2 score.
- Other Modeling Techniques:
  - Polynomial Features: A pre-processing step that generates polynomial and interaction features.
- Numerical Computation:
  - NumPy[36]: A library for numerical computing with support for large, multi-dimensional arrays and matrices.

Specifically for the project's Greek validation sites (owned by partners PPC and PPCR), the development and execution of the WFFM will utilise Matlab[37] for applying autoregressive integrated moving average (ARIMA), and Artificial Neural Network (ANN) models. Thus, the technology stack for these validation sites includes Matlab commands and functions that allows data processing, model development, training, and visualisation.

32 https://xgboost.readthedocs.io/en/stable/#
33 https://joblib.readthedocs.io/en/stable/
34 https://matplotlib.org/stable/
35 https://seaborn.pydata.org
36 https://numpy.org
37 https://www.mathworks.com/products/matlab.html

4.4 Power Prediction Module: functional and technical specifications

The functional and technical specifications of the Power Prediction Module (PPM) are still under development due to the workflow phases in WP6 and WP7. It is necessary to complete the water flow/volume forecast, before calculating the available power prediction. The PPM be described in detail in the associated WP6 deliverable, D6.2 (due in M18); nevertheless, a preliminary architecture is proposed within this deliverable.

The PPM will utilise the output from the Weather and Flow Forecasting Module. This data, combined with historical energy production data, historical water level data, and historical turbined water data, will serve as input for the module. Using these historical variables, the system's efficiency will be calculated through a regressor and utilised in the power calculation module in WP6 and WP7. The output of the power prediction module will be provided in various graphical formats, such as PNG, JPEG, or SVG, Additionally, a .csv file or a dataframe will be generated.

The following figure illustrates a schematic representation of the general architecture for the Power Prediction module.

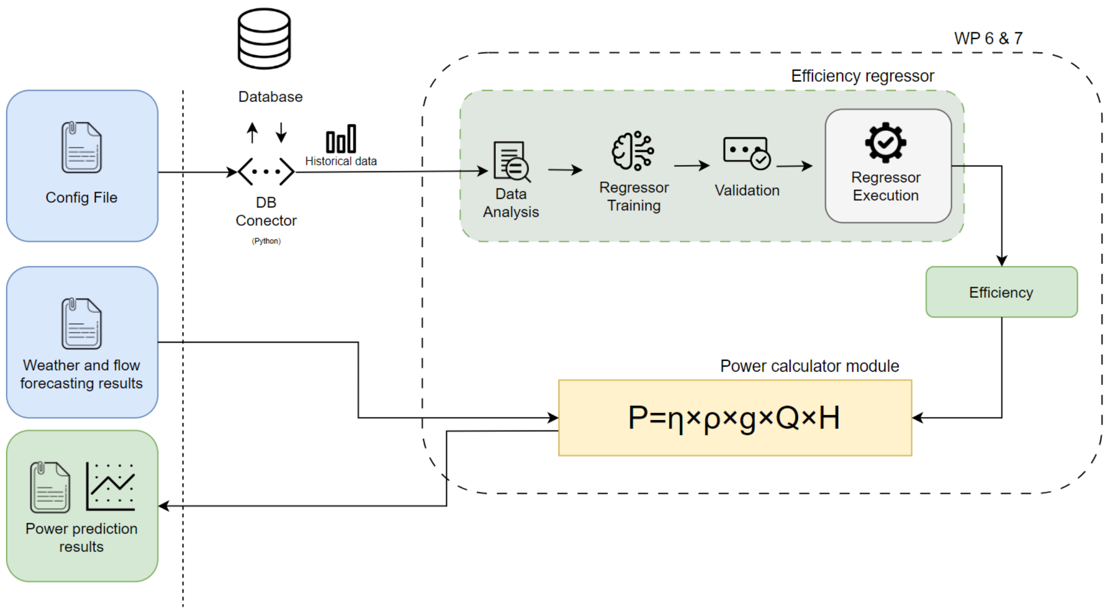

Figure 9 Power Prediction Module's internal architecture

4.4.1 Power Prediction Module: Subcomponents description

The key components of the Power Prediction Module (PPM) are described in the following table along with their mapping to the iAMP-Hydro UCs and associated requirements (defined in Chapter 2 and 3 respectively).

Table 32 Weather & Flow Forecasting Module*: Subcomponents description and mappings to iAMP-Hydro UCs and requirements*

Comp. ID	Sub-component Name	Subcomponent Description	Related UC(s)	Related Reqs. (Chapter 3)
PPM	Config file	Component designed to manage the operations and inputs required for the weather and flow forecasting module as well as the power prediction module.	-	-
PPM		Handles the connection with the iDML to retrieve historical data, including flow/volume data, climate	UC_6_2	Req_017, Req_018

	Database connector	variables (precipitation, temperature), historical energy production data, water level data, and turbined water data. This connection is essential for accessing the input data necessary for model execution.		
PPM	Efficiency regressor	The efficiency regressor module is designed to infer the total system efficiency as a function of gross head and flow, using historical data samples. This module utilizes regression techniques to predict efficiency.	UC_6_2	Req_016, Req_019
PPM	Power calculator module	Model encapsulated in .pkl files or other equivalent formats. Model execution outputs the available power prediction.	UC_6_2	Req_016, Req_019

Technology stack

The technology stack for the Power Prediction Module is currently under development. However, it is anticipated that this module will utilise similar technologies to those employed in the Weather and Flow Forecasting Module. The key components that will be shared across both modules for CUERVA Validation Sites include:

- Data Processing and Management:
  - Pandas[5]: Used for data manipulation and analysis, providing data structures for handling datasets.
- Data Visualisation:
  - Matplotlib[34]: A plotting library used for visualizations in Python.
  - Seaborn[35]: A data visualization library based on Matplotlib.
- Numerical Computation:
  - NumPy[36]: A library for numerical computing with support for large, multi-dimensional arrays and matrices.

All the key components developed across both modules, presented in Table 31 above will be applied also for PPC and PPCR Validation Sites.

5 Conclusions and next steps

The document at hand, D9.1 - Platform architecture designed and released, outlines the key results of T9.1 and T9.2 activities, which are of great importance in setting the basis for the iAMP-Hydro's successful implementation. Starting by defining a comprehensive process of defining the project's use cases and extracting the associated requirements, D9.1 establishes the basis for the design and development of the iAMP-Hydro ICT infrastructure. Emphasis has been given on ensuring that the system architecture meets both technical and practical needs, driven by the active collaboration between the (first) end-users, being the project's pilots and external stakeholders.

In this direction, D9.1 provides the definition of iAMP-Hydro UCs in Chapter 2, enabling the identification of specific functional needs of the hydro-energy sector; and which in turn facilitated the elicitation of the iAMP-Hydro requirements, as documented in Chapter 3. Following, the requirements were translated into a robust architectural framework, ensuring that iAMP-Hydro ICT Infrastructure clearly addresses challenges related to data management, decision optimisation, weather and flow forecasting, predictive maintenance, and power prediction. Chapter 4 presents the detailed technical and functional specifications of the key components of the iAMP platform, specifically the intelligent Data Management Layer and the integrated Decision Optimization Layer. Furthermore, other components of the iAMP-Hydro ICT Infrastructure such as the Predictive Maintenance Module, the Weather & Flow Forecasting Module, and the Power Prediction Module were thoroughly discussed, highlighting their importance in achieving the project's objectives of transforming the hydro-energy sector through data-driven solutions.

As the iAMP-Hydro project progresses, the following essential steps are anticipated in further refining and implementing the iAMP-Hydro ICT Infrastructure:

- Release of the first version of iAMP-Hydro Platform (M20) that will be documented in D10.1 providing a more detailed breakdown of the platform's architecture and its initial functionalities. In the first version of the iAMP-Hydro Platform, it is anticipated that the majority of its components will be integrated and will be utilised by the project's demonstrators to validate their operational effectiveness and provide feedback.
- Enhance stakeholder engagement; considering that the success of the iAMP-Hydro ICT infrastructure is highly dependent on its usability and relevance to hydro-energy sector end-users, continued engagement with stakeholders will be sought, including feedback gathering from both the project's partners and external users (through the active end-user participation and involvement as part of the COP and co-development workshops undertaken in the context of WP13-WP14), towards ensuring it remains user-friendly and addresses evolving industry requirements that can be addressed in the platform's final release (M28).
- Testing and validation of the various components outlined herein; these will undergo testing and validation to ensure they meet the specified requirements; including the utilisation of the first release of the iAMP-Hydro platform, to be used as a testbed for the iAMP-Hydro Demonstrators' activities.
- Iterative updates, based on feedback from testing and further requirements gathering, also including updates on both the technical architecture and the platform's functionalities, if needed. This iterative process will ensure the iAMP-Hydro platform remains aligned with the project's goals and can accommodate future technological advancements in the platform's final release due in M28.

References

Description of Action - Ref. Ares(2024)4162214 - 10/06/2024). intelligent Asset Management Platform for Hydropower operation and maintenance (iAMP-Hydro)

IEC 62559-2:2015, Use case methodology - Part 2: Definition of the templates for use cases, actor list and requirements list