D8.1 – Hydro and energy data modelling landscape analysis

March 2024

Suite5

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

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	23/02/2024	Draft ToC
0.20	27/02/2024	Assigned partners roles and sent for contributions
0.30	13/03/2024	Updated as per partners contributions
0.60	20/03/2024	Updated as per internal review comments and sent for peer review
0.90	28/03/2024	Updated as per peer review comments
1.00	29/03/2024	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

In accordance with the DoA, the present deliverable D8.1, is an output of T8.1 activities which aims scrutinize the landscape of hydro and energy data modeling, with a focus on identifying open standards, semantic models, and ontologies that align with the iAMP-Hydro project's scope. This deliverable conduct a thorough examination and evaluation of well-established standards, ontologies, and data models pertinent to the Hydro-Energy domain, while assessing their suitability to fulfil the data requirements of the iAMP-Hydro project, ensuring they address the specific needs while also identifying any gaps in data modelling that may exist. Furthermore, D8.1 initiates the preliminary identification and definition of high-level entities and their respective domain concepts, based on the project's semantic data needs and which are intended to serve as the foundational elements of the common information data model, to be leveraged for the purposes of the iDML. This assessment lays the groundwork for developing a comprehensive data model tailored to the unique requirements of the iAMP-Hydro, facilitating informed decision-making and strategic planning moving forward.

Table of Contents

List of Figures

List of Tables

List of Acronyms

Acronym	Full description
AI	Artificial Intelligence
DL	Deep Learning
HPP(s)	Hydro power plant(s)
iAMP	intelligent Asset Management Platform
iDML	intelligent Data Management Layer
iDOL	intelligent Data-driven Optimization Layer
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Standards Organisation
ML	Machine Learning
Mx	Month x
O&M	Operation and Maintenance
RES	Renewable Energy Sources/Systems
Tx.y	Task x.y
WPx	Work Package x

1 Introduction

1.1 Description of the deliverable content and purpose

The present deliverable, namely D8.1 - Hydro and energy data modelling landscape analysis, reports the activities undertaken in the context of T8.1: Analysis of Existing Standards-based Data Models and Definition of Modelling Needs.

In accordance with the DoA [1], the objective of this task is to identify and assess the current landscape of hydro and energy data modelling; and specifically select open standards, semantic models, and ontologies based on their relevance to be included in the iAMP-Hydro project's, iDML data model that will be clearly defined in D8.2 - iDML Life-cycle Management Design, due in M15.

In pursuit of these objectives, a comprehensive analysis has been carried out assessing current standards ontologies and data models in terms of their ability to meet the project's data requirements, evaluating their completeness, and identifying any notable gaps. Such a methodical approach ensures a meticulous selection process, aligning the chosen models and standards closely with the project's overarching needs and objectives.

Furthermore, T8.1 delves into the examination of relevant to the hydro-energy industry modeling paradigms, alongside the analysis of data assets present at the project's validation sites (at the time of writing this deliverable) and those that will be needed in the future. The intent here is to extract the foundational High level Information Entities and Domain Concepts to be supported by the iDML while also exploring alternative data modeling approaches for concepts that are not sufficiently covered by existing models, and ontologies. This work is critical for laying the groundwork for subsequent developments in the project's data modeling efforts and ensuring a robust and comprehensive approach for the integration of data within the project's scope.

1.2 Structure of the document

To fulfil the objectives of T8.1 the remainder of this deliverable is structured as follows:

- Chapter 2 outlines the theoretical and practical methodologies employed to gather, assess, and combine the relevant data and standards; setting the foundation for the investigative framework used throughout D8.1.
- Chapter 3 identifies the specific goals, objectives, and requirements integral to the iAMP-Hydro project, encompassing an initial identification of the project's generic data requirements and analysis of the semantic scope of the iDML, towards extracting the relevant high level information entities and domain concepts that can be included in its data model.
- Chapter 4 presents a comprehensive literature review of the current landscape in hydro and energy data modelling, by undertaking a thorough review of existing standards categorized by their issuing bodies, relevant ontologies and data models.
- Chapter 5 integrates the findings from the previous chapters, identifying existing gaps in the current hydro and energy data modeling landscape and extracting the main takeaways specific to the iDML.

Finally in Chapter 6 we conclude the document by summarising the key findings and insights gained throughout the deliverable.

2 Methodology

The initial step in examining known standards, ontologies data models for their relevance in the iAMP-Hydro project involved creating a preliminary list of such material currently utilised in the hydro power and energy sector in general, along with the domains of: a) condition monitoring and predictive maintenance, b) ecological status monitoring and water management, c) improved weather and flow forecasting; where the iAMP-Hydro project is expected to produce innovative digital solutions.

To achieve this, a literature review was conducted in the specified domains, and specific standards, ontologies, and data models (see Chapter 4) were chosen without consideration of their practical applicability. In summary our review focused on identifying and comprehending as relevant to the project's scope 26 standards, 6 ontologies and 5 data models which are presented in more detail in Chapter 4.

In an effort to narrow down the review to the research scope of the project, we engaged in a collaborative process with both technical and demonstration partners from the consortium. The aim was to refine and narrow the initial list of standards, ontologies, and data models, ensuring they align with the specific needs of the hydropower sector and the project. Partners' insights were actively sought and their feedback was collected to assess the practicality and relevance of each standard, ontology, and model in real-world settings. Concerns regarding the applicability of certain frameworks were also collected, helping to prioritise those most commonly used in our partners' daily operations for deeper analysis and potential integration into the iAMP-Hydro project. Towards systematically gathering this critical feedback, Table 3 presented below was distributed to all consortium members. Each partner was requested to evaluate and rank each standard, ontology, and data model based on its effectiveness and applicability to both the overarching goals of the project and their individual operational requirements. This structured approach ensures that the collaborative efforts are grounded in practicality and geared towards enhancing the efficiency and sustainability of hydropower operations.

To assess the effectiveness of the various data models, standards and ontologies these were methodically categorised based on their relevance and applicability to the projects key results (see Table 3) pertaining to:

- Data-driven operation and maintenance
- Weather & Flow Forecasting
- Condition Monitoring & Predictive Maintenance
- Biodiversity & ecological status monitoring
- Data gathering, secure communication and standardisation.

In line with this classification, a thorough evaluation of each standard, ontology, and data model was conducted to pinpoint their relevance to the overarching iAMP-Hydro goals, while identifying any potential constraints or limitations. Furthermore, given the active utilisation of these data models in the operations of the project's demonstration partners and their widespread adoption across the hydro-energy sector, their feedback was sought. This feedback aimed to identify and prioritise the most impactful standards, ontologies, and data models that are also pertinent to their operational needs. Their feedback was then incorporated, facilitating the selection of the most suitable standards, ontologies, and data models that align with the strategic objectives of the iAMP-Hydro ensuring that the chosen frameworks are not only theoretically sound but also practically beneficial and applicable to the needs of the industry.

The results of this analysis approach led to the identification of the most appropriate suitable standards/ontologies/data models for the purposes of iAMP-Hydro. In addition, our approach entails the provision of a gap analysis (in chapter 5) between the ontologies and data model where varying levels of maturity among them have been recognised and chosen according to the specific focus of iAMP-Hydro project.

According to our analysis, the scope of the selected standards/ontologies/data models covers a range of critical aspects in Hydro Power plants (HPPs) monitoring and optimization, encompassing operational and sensor data, performance metrics, renewable energy forecasts, etc. towards ensuring a holistic approach to data analysis for informed decision-making in the hydro power generation process.

Table 3. Hydropower sector's Standards, Ontologies, and Data Models - Applicability Ranking

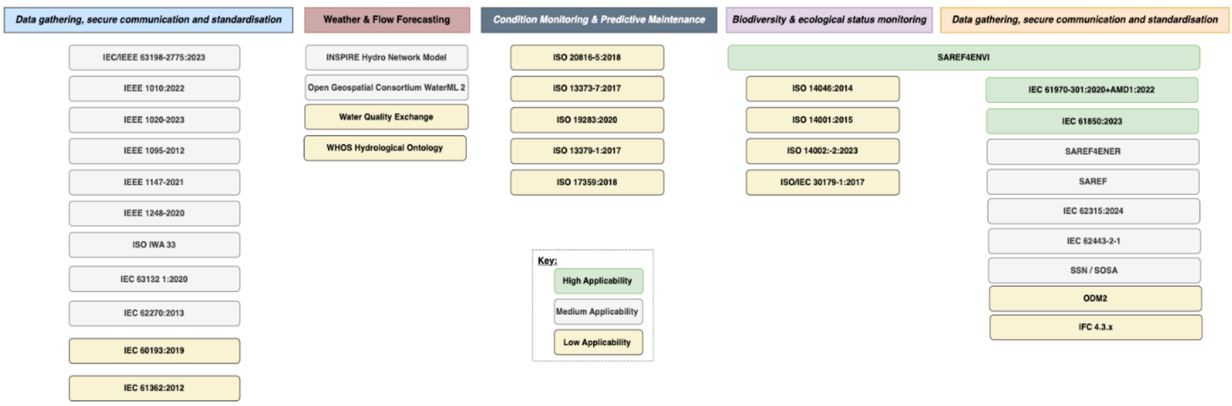
Standard	**Data Model/ Ontology availability**	**Title**	**Aggregated Ranking applicability** **1 -> Low** **2 -> Medium** **3 -> High**
Data-driven operation and maintenance			
IEC/IEEE 63198-2775:2023		Technical Guidelines for Smart Hydroelectric Power Plant	**2**
IEEE 1010:2022		IEEE Guide for Control of Hydroelectric Power Plants	**2**
IEEE 1020-2023		IEEE Guide for Control of Small (100 kVA to 5 MVA) Hydroelectric Power Plants	**2**
IEEE 1095-2012		IEEE Guide for the Installation of Vertical Generators and Generator/Motors for Hydroelectric Applications	**2**
IEEE 1147-2021		IEEE Guide for the Rehabilitation of Hydroelectric Power Plants	**2**
IEEE 1248-2020		IEEE Guide for the Commissioning of Electrical Systems in Hydroelectric Power Plants	**2**
ISO IWA 33		Technical guidelines for the development of small hydropower plants	**2**
IEC 63132 1:2020		Guidance for installation procedures and tolerances of hydroelectric machines - Part 1: General aspects	**2**
IEC 62270:2013		Guide for Computer-Based Control for Hydroelectric Power Plant Automation	**2**
IEC 60193:2019		Hydraulic turbines, storage pumps and pump-turbines - Model acceptance tests	**2**
IEC 61362:2012		Guide To Specification of Hydraulic Turbine Governing Systems	**2**
Weather & Flow Forecasting			
WHOS	✓	WHOS Hydrological Ontology	**1**
INSPIRE	✓	INSPIRE Hydro Network Model	**2**
OGC WaterML2	✓	Open Geospatial Consortium WaterML 2	**2**
WQX	✓	Water Quality Exchange	**1**
Condition Monitoring & Predictive Maintenance			

ISO 20816-5:2018		Mechanical vibration - Measurement and evaluation of machine vibration - Part 5: Machine sets in hydraulic power generating and pump-storage plants	**1**
ISO 13373-7:2017		Condition Monitoring and Diagnostics of Machines - Vibration Condition Monitoring - Part 7: Diagnostic Techniques for Machine Sets in Hydraulic Power Generating and Pump-Storage Plants	**1**
ISO 19283:2020		Condition monitoring and diagnostics of machines | Hydroelectric generating units"	**1**
ISO 13379-1:2017		Condition monitoring and diagnostics of machines - Data interpretation and diagnostics techniques - Part 1: General guidelines	**1**
ISO 17359:2018		Condition monitoring and diagnostics of machines - General guidelines	**1**
Biodiversity & ecological status monitoring			
SAREF4ENVI	✓	SAREF for the environment	**3**
ISO 14046:2014		Environmental management - Water footprint - Principles, requirements and guidelines	**2**
ISO 14001:2015		Environmental management systems - Requirements with guidance for use	**2**
ISO 14002:-2:2023		Environmental management systems Guidelines for using ISO 14001 to address environmental aspects and conditions within an environmental topic area - Part 2: Water	**2**
ISO/IEC 30179-1:2017		Internet of Things (IoT) - Overview and general requirements of IoT system for ecological environment monitoring	**2**
Data gathering, secure communication and standardisation			
IEC 61970-301:2020+AMD1 :2022 CSV	✓	Energy management system application program interface (EMS-API) - Part 301: Common information model (CIM) base Common Information Model	**3**
IEC 61850:2023	✓	Communication networks and systems for power utility automation	**3**
IEC 62315:2024		Power systems management and associated information exchange - Data and communications security - ALL PARTS	**1**
IEC 62443-2-1		Industrial communication networks - Network and system security - Part 2-1: Establishing an industrial automation and control system security program	**1**
SAREF	✓	Smart Applications REFerence ontology	**2**
SSN-SOSA	✓	Semantic Sensor Network - Sensor, Observation, Sample, and Actuator	**2**
SAREF4ENER	✓	SAREF for the energy system	**2**

IFC 4.3.x	✓	Industry Foundation Classes	**1**
ODM2	✓	Observations Data Model 2	**2**
IEC 62325	✓	Framework for energy market communications - Part 301: Common information model (CIM) extensions for markets	**1**
IEC 61968-100:2022		Application integration at electric utilities - System interfaces for distribution management - Part 100: IEC implementation profiles for application integration	**1**

Figure 1 illustrates the categorisation and coverage of the evaluated standards, data models, and ontologies, organised according to the principal verticals (main interests) of the project. Additionally, the varying colours denote the aggregated rankings displayed in the preceding table. This visualisation aids in understanding the application scope and the integration potential of each standard within iAMP-Hydro's framework.

Figure 1. Categorisation and coverage of standards, data models and ontologies reviewed for the iAMP-Hydro needs



3 Identifying the data requirements of iAMP-Hydro

Before delving into identifying different standards, ontologies, and data models towards extracting crucial information entities relevant in the hydro-energy field, that can be mapped to the principal iAMP-Hydro semantic data requirements; this section presents a concise overview of specific data specifications steaming from the project's objectives, supplemented by further data specifications extracted from an initial data landscaping exercise. In this context, the following sections aim to define key data specifications and extract the High-level Information entities and relevant domain concepts that are aligned with the project's objectives, and which will support the development of a targeted data model tailored for the iAMP-Hydro. This overview addresses the requirements of both the demonstration partners (hydro operators) and the technical partners involved in executing the project's activities.

The iAMP-Hydro project aims to enhance the digital operations of current hydro-power plants by developing six innovative results, which collectively will constitute the intelligent Asset Management Platform (iAMP), a highly effective integrated framework (built on state-of-the art technologies for data sharing, real-time data analytics as well as data lifecycle management) for introducing additional value and delivering significant benefits to the hydropower sector stakeholders incorporating secure, open, and transparent data-sharing protocol. More specifically, the iAM Platform is structured into two interconnected but distinct layers:

i. The *Intelligent Data Management Layer (iDML), s*erving as the hub for gathering and storing hydro-related data. It is equipped with a data analytics engine that supports the training of machine learning (ML) and deep learning (DL) models. Additionally, it allows for the integration of existing, pre-trained AI models, simplifying the process for hydro facility operators.
ii. The *Intelligent Data-driven Optimization Layer* (iDOL), which utilises the insights generated by the newly developed models for predictive maintenance, ecological status assessment, and weather and flow prediction. These insights are then combined with additional external data sources, such as renewable energy systems (RES) and energy market information, to craft data-informed strategies aimed at improving the operation and maintenance (O&M) of hydro facilities.

Moreover, the iAMP-Hydro project will deploy a set of innovative sensors towards providing input to the iAMP Management Platform, which in turn will enable the measuring and integration of several indicators providing critical information on electromechanical equipment condition (vibration, acoustics, temperature & electrical), ecological status and biodiversity (biological, physical & chemical water parameters), and weather and flow forecasts.

Overall, the objective of the iAMP is to support hydropower operators towards improving predictability, flexibility, and sustainability in their operations and to reduce failures and O&M costs of their assets.

Within this framework, the generic data specifications that are of interest to the iAMP-Hydro, derive from the different project results and activities, pertaining to:

- *Condition Monitoring & Predictive Maintenance*: Here it expected that sensors fitted on equipment will feed data (vibration, acoustics, temperature, voltage and current) in the iDML to be stored, harmonised, etc., and ultimately utilised for analysis to detect indications of hydromechanical and electrical equipment failure mechanisms, such as cavitation, bearing, wear, etc. Laboratory turbine testing will also take place utilising data on flow, pressure, torque, and rpm, along with Computational Fluid Dynamics (CFD) modelling to simulate failures of different machine sizes and differing operating conditions.
- *Ecological Status Monitoring and Water Management*: In this instance, a novel sensor with combined digital probes for temperature, water depth and gas supersaturation will be developed, and a data logger will be adapted for the digital probe, and cloud interaction. Data on site-specific local flow regimes and local hydrological information temperature and flood regime regulations will

be leveraged towards enabling ecological optimisations and enhancing (where feasible) water availability for energy production by adjusting for local flow regimes and hydrological conditions.

- *Improved Weather and Flow Forecasting*: Here a novel weather and flow forecasting analytics tool will be implemented (utilising advanced ML algorithms that needs to be trained and adapted to a variety of use cases, operation conditions and climate regions), while also leveraging diverse open data sources (e.g. satellite data, GIS data, power production, weather forecast...).

Leveraging all this information, Table 4 below outlines the core high-level information entities and a non-exhausting list of domain concepts relevant to the iAMP-Hydro, deriving from the project's innovative results and aligning with its overarching objective.

Table 4. Breakdown of f High-level entities and their domain concepts as extracted from the project's objectives

High-level Information entities	Domain concepts
Condition Monitoring Data	Sensor, Measurement, Equipment, Hydro Plant, Electrical Equipment, Failure, Operational Conditions, Location
Predictive Maintenance Data	Hydro Plant, Power Station, Equipment Operation, Location, Failure, Maintenance Action, Maintenance Event, Maintenance Plan, Sensor, Measurement, Supplier, Operational Conditions
Ecological Status Monitoring Data	Water Measurements, Biological Parameters, Chemical Parameters, Sensor, Measurement, Hydro Plant, Ecological Conditions, Biodiversity Conditions, location
Water Management Data	Reservoir, Gas Supersaturation, Hydrological Measurement, Sensor, Measurement, Hydro Plant, Water Measurements, Energy Production, Ecological Conditions, location
Weather Forecasting Data	Weather Conditions, Location, Sensor, Hydro Plant, Forecasting Models, Operational Conditions
Flow Forecasting Data	Water Flow, Forecasting Models, Sensor, Hydro Plant, Operational Conditions, Location, Energy Production

3.1 Data requirements arising from the data landscaping exercise

In this section, we outline the initial data needs identified by the project's partners perspective, encompassing hydro-operators (representatives of the project's pilot sites) and technical partners tasked with developing advanced sensors, forecasting algorithms, all of which are to be integrated within the iAM Platform. These entities represent the primary stakeholders and ultimate beneficiaries of the platform. Following a comprehensive initial data landscaping exercise at the early stages of the project's, we now define the data presently available and the data necessary for the stakeholders' effective functioning within the iAMP-Hydro project framework.

During the data landscaping exercise, all partners were asked to provide info regarding their available data sets (that will be used for the project's purposes), along with the data required towards efficiently undertaking their tasks, such as construction of mathematical models, the creation of forecasting algorithms, etc. In this direction an appropriate Excel template was circulated among the partners, providing a structured approach to cataloging, and understanding the available and required data assets for iAMP-Hydro. This comprehensive approach ensures also that all data assets are properly documented, maintained, and utilised in a manner that respects owner's privacy and ultimately enhancing the efficiency and effectiveness of the project's data management practices.

The Excel template is designed to meticulously gather fundamental details of available/required data, such as the dataset's unique identifier, title, description, and key concepts included within each data set. It extends to detailed features of the data, including its volume, variety, type, format, and how rapidly it is generated, as well as its historical availability, temporal and spatial coverage, language, adherence to relevant standards, and temporal resolution. Additionally, the availability of the data assets recorded is asked, also detailing the methods by which they can be accessed, the frequency of updates, and the strategy for keeping the data up-to date. Finally, the template addresses data rights, prompting partners to classify their recorded data as Private, Confidential, or Public, ensuring clear understanding and proper handling of data privacy. It shall be noted that this Excel document will be regularly updated throughout the duration of the project to accurately document and track all emerging data assets required by our partners for the efficient execution of their tasks. This dynamic approach ensures that we maintain a comprehensive and current overview of all necessary data, facilitating effective collaboration and task management.

Having provided the description of the Excel template, we summarise below the various data specifications provided from the different demo sites (Spanish and Greek sites) and technical partners, focusing on extracting potential High-level Information Entities and Domain concepts suitable for the project's needs and which eventually will be integrated into the iDML.

Starting with the Spanish demos, the data assets provided, encompass production and flow measurements (timestamp and measurement value) for these three small HPP pilots. Specifically, for one specific demo additional data are available, such as water input/output, levels, precipitation, temperature, and volume.

Regarding the Greek demo, there are available both monthly and daily reports that include energy metering and water flow data. This encompasses detailed information for generator units and transformers, such as voltage, current, power metrics, energy production, and more. Also available are data pertaining to the reservoir, covering aspects like water level, reservoir reserve, overflow from reservoir's spillway and station's flap gates, water consumption for production, special consumption, performance grade, ecological flow, water consumption flow, irrigation water flow, reservoir inflow, losses. From the technical partners' perspective, there are both laboratory and Computational Fluid Dynamics (CFD) data available, including parameters such as vibration, acoustics, pressure pulsations, power, and temperature across various conditions. These data sets also cover efficiency, cavitation, and energy prices from different countries, along with forecasts for Renewable Energy Sources (RES), specifically PV and Wind Energy.

In addition, technical partners also provided their minimum data requirement towards efficiently undertaking their operations pertaining to the development of advanced ecological sensors, predictive maintenance models along with flow and power prediction forecasting models. The required data shall encompass measurements of total inflow into the reservoir, power or energy production data, reservoir level and reservoir free surface level. Additionally, critical environmental parameters like precipitation and air temperature are necessary to ensure comprehensive and accurate forecasting.

It's important to recognise that the data provided by the partners serve as an initial framework for identifying key information entities and associated domain concepts for integration into the iDML common information model (CIM). However, the scope of this common information model (or data model) is anticipated to expand throughout the project's duration, incorporating new concepts to meet the evolving needs of project stakeholders. The information supplied thus far has enabled an initial evaluation of the common information model's specifications tailored to the various stakeholders involved.

Having presented above the existing and required data sets (at the time of writing this deliverable), as extracted through the data landscaping activity, we present in Table 5 a breakdown of additional high level Information entities and domain concepts.

Table 5. Breakdown of High-level entities and their domain concepts as extracted from the data landscaping activity

High-level Information entities	Domain concepts
Hydro Plant Energy Demand Data	Hydro Plant, Power Station, Measurement, Sensor, Equipment, Location, Electricity Meter, Hydro Operator, Generator, Transformer, AC Line, DC Line, Period
Hydro Plant Energy Production Data	Hydro Plant, Power Station, Measurement, Sensor, Equipment, Location, Hydro Operator, Grid, Alternator, Turbine, Transformer, AC Line, DC Line, Period
Reservoir Data	Reservoir Characteristics, Dam Characteristics, Sensor, Measurement, Location
Water Flow Measurement Data	Water Flow, Sensor, Measurement, Location, Period
Water Flow Forecast Data	Flow, Sensor, Measurement, Location
Weather Data	Weather Station, Weather Measurement, Sensor, Location
Weather Forecast Data	Weather Station, Weather Measurement, Sensor, Location
RES Generation Forecast Data	RES Asset Profile Data, RES Generation System (PV, Wind, etc.), Measurement, Period
Energy Pricing Data	Customer, Energy Market, Pricing Structure, Tariff Profile
Sensor/Field Measurements Data	Sensor, Measurement, Location

While the above sections provide a preliminary step towards defining the iDML's High level Information entities and domain concepts, it is important to delve deeper into their specifics to define further data concepts involved; in this direction, a thorough examination is scheduled to be undertaken as part of T8.2 activities (and documented in D8.2, due in M15) towards extracting the nominal or configuration parameters and their measurements, alongside more complex entities shaped by numerous concepts based on the analysis of both existing data and future data needs from the demo and technical partners.

3.2 Analysis of the iAMP-Hydro Data Management Layer (iDML) semantic scope

As defined in the DoA [1], iAMP-Hydro will deliver a comprehensive standards-based data management layer, namely the iAMP-Hydro Data Management Layer (iDML) of the iAM Platform. To effectively fulfil its operations, the iDML will leverage a common information model that shall be based on existing standards, ontologies and data models and be properly enriched and extended with new concepts and attributes tailored to the data availability and needs of the iAMP-Hydro project, towards effectively facilitating common data integration and sharing practices. The core benefits of having the iDML's common data model that can be shared among hydro operator and relevant vendors it that it will provide them with strategic independence by promoting a standard alignment, improved performance from its unique naming conventions and thus facilitate higher levels of interoperability among hydro-operators and vendors.

While the primary objective of iAMP-Hydro project is to enhance and support the digital transformation of existing hydropower facilities, the iDML's data model shall be designed to have a more extensive reach. This broader perspective aligns with the project's overarching goals, which include the development of the iAM Platform. Although this platform will be specifically designed for the hydropower sector, its utility extends beyond it, offering a suite of solutions that include secure, open, and transparent data-sharing protocols. Additionally, it integrates three innovative digital solutions focused on monitoring and predictive maintenance, ecological status and water management, and advanced weather and flow predictions, illustrating the project's comprehensive approach and wider applicability.

Furthermore, it is essential to recognize that, like any significant infrastructure, hydropower plants generate vast amounts of data throughout their lifecycle, from initial planning stages to operational status. These facilities interact with a variety of systems, requiring detailed data models to encapsulate critical data semantics effectively. Consequently, such a digitalization process should cover a range of elements beyond plain hydro power generation. This includes integrating sensor networks, supervisory control and data acquisition (SCADA) systems, communication protocols, distributed energy resources, energy storage solutions, and environmental monitoring via external sensors. Therefore, the iDML needs to be flexible enough to meet the complex semantic needs of hydropower operations and the broader energy landscape.

In addition to its comprehensive scope, the iAMP-Hydro must address the challenge posed by the uneven rates of data generation within hydropower plants. This irregularity often is presented in the form of varying spatiotemporal data resolutions, leading to discrepancies where sensor data does not correspond directly with the growing volumes of data over time. Consequently, it is crucial for the iAM Platform to effectively handle large quantities of raw data without resorting to unnecessary discarding or irrational aggregation. Therefore, the iDML should be optimised to minimise any excessive computational demands, ensuring it aligns seamlessly with the big data requirements utilised within the hydro power domain.

Another crucial feature of the iDML lies in its flexibility and extensibility. It is not reasonable to anticipate that, upon publishing its data model, it will be considered as definitive and applicable to all future needs of the project. On the contrary, throughout the life cycle of the iAMP-Hydro project we expect emergence of new data modelling needs, as well as modifications and updates to the existing data model. Hence, mechanisms must be in place to facilitate the upgrading and enrichment of this common data model.

Consequently, the fundamental characteristics imperative for the common information model of the iDML are presented below:

- It shall encompass all relevant data semantics associated with the operation and verticals of the hydropower-energy domain (also covering all the iAMP-Hydro project's needs), ensuring comprehensive coverage, and integrating interoperable links with external standards and recognised models currently in use.
- It shall be capable of handling large datasets typical of big data environments, ensuring it delivers high performance at a scale appropriate for big data applications.
- It shall include mechanisms for life cycle management that allow for its expansion, showcasing its ability to adapt over time.

4 Hydro and energy data modelling landscape - Literature Review

In this chapter, we conduct a comprehensive evaluation of the existing standards, ontologies, and data models identified in literature, which are relevant to the hydropower sector and align closely with the scope of the iDML. In this context the following sections delve into their relevance and applicability concerning the scope of the iAMP-Hydro project, taking also into account the crucial requirements outlined in the previous chapter. This assessment aims to enrich our understanding and ensure that the chosen frameworks for the iDML's semantic data model effectively support the project's objectives, while providing a structured overview of existing guidelines and their impact on the project's scope.

4.1 Review of Existing Standards

4.1.1 IEEE /IEC standards

The Institute of Electrical and Electronics Engineers (IEEE) and the International Electrotechnical Commission have established a comprehensive framework of standards fundamental to the advancement and harmonization of electrical and electronic engineering disciplines. Many of the standards, are considered as integral to the iAMP-Hydro project, particularly in areas such as electrical power generation, distribution, telecommunications, and Information Technology. The latter covering a broad spectrum, including network infrastructure, communications, data exchange protocols, cybersecurity, and information processing systems. Additionally, the hydroelectric energy sector benefits from a specific range of IEEE standards guiding various aspects such as unique identification in facilities, control systems for power plants, and computer-based control for automation. Additionally, there are standards for small hydroelectric power plants and rehabilitation of existing plants, ensuring safety, efficiency, and technological updates. These standards are crucial for the digitalisation and effective operation of hydroelectric power generation systems and are also reported.

4.1.1.1 IEC 60193:2019

The IEC 60193:2019, titled “Hydraulic turbines, storage pumps and pump-turbines - Model acceptance tests” applies to laboratory models of any type of impulse or reaction hydraulic turbine, storage pump or pump-turbine. It applies to models of prototype machines either with unit power greater than 5 MW or with reference diameter greater than 3 m. This document excludes all matters of purely of commercial interest, except those inextricably bound up with the conduct of the tests. It covers the arrangements for model acceptance tests to be performed on hydraulic turbines, storage pumps and pump-turbines to determine if the main hydraulic performance contract guarantees have been satisfied. It contains the rules governing test conduct and prescribes measures to be taken if any phase of the tests is disputed; and outlines procedures for acceptance tests on hydraulic turbines, storage pumps, and pump-turbines to verify compliance with main hydraulic performance contract guarantees. It defines terms and quantities, specifies testing and measurement methods, computation procedures, and defines report requirements. Guarantees may be based on either prototype or model hydraulic performance. This edition incorporates technical updates such as modernized methods and tools for dimensional checks, improved accuracy requirements, streamlined sections, updated measurement techniques, revised terminology, and enhanced testing methodologies. Key areas of update include dimensional checks, discharge measurement, pressure fluctuation analysis, surface requirements, cavitation nuclei content, radial thrust measurements, hydraulic loads on control components, extended operating range testing, index testing, roughness measurement methods, references, figures, sigma definition, and transposition methods in line with IEC 62097.

4.1.1.2 IEC 61850:2023

The IEC 61850:2023, Published by IEC Technical Committee 57 and titled “Communication networks and systems for power utility automation“ is an international standard widely utilised for the design and operation of substation automation systems, extensively applied in the power utility industry. It provides a comprehensive framework for the communication, integration and interoperability of various devices and systems within a substation [2]. While in its first release the focus was on substation communication networks and system management, presently it applicable goes beyond utilities, widely exploited in industries such as wind energy, EV charging, while its principles can also be applied to hydroelectric power plants. The emergence of IEC 61850 stemmed from the necessity to consolidate the various protocols employed by organizations, towards facilitating communication and maintaining the functionality of their network infrastructures. This unification is particularly challenging for mission-critical organizations, where its crucial to not only ensure seamless communication but also to maintain the security of their infrastructures. IEC 61850 not only addresses the need for unified communication but also emphasizes interoperability, a fundamental concept enabling the integration and management of equipment sourced from diverse manufacturers. In the contemporary landscape, the global trend is undeniably leaning towards the widespread adoption of the IEC 61850 standard.

More specifically, one of these standards, IEC 61850-7-410:2012 "Communication networks and systems for power utility automation - Part 7-410: Basic communication structure - Hydroelectric power plants - Communication for monitoring and control", deals specifically with hydro power and specifies the additional common data classes, logical nodes and data objects required for the use of IEC 61850 in a hydropower plant [2].

Overall, the IEC 61850 offers a robust framework for communication, data modeling, interoperability, and cybersecurity in HPP Its adoption can lead to more efficient and reliable operations, streamlined maintenance processes, and better integration with the broader power system. The IEC 61850 defines a standard communication protocol for power system devices within a substation. As such, the standard's communication capabilities are highly relevant to HPPs and can be utilised to enable seamless communication and information exchange between various devices and control systems within the plant. In terms of Data Modelling the IEC 61850 employs a standardised object-oriented data modeling approach, allowing for consistent representation of data across different devices and systems. In HPPs, this standardized data modeling ensures uniformity in representing information related to turbines, generators, sensors, and other equipment. It facilitates easy integration and interoperability. The IEC 61850 emphasises on interoperability between devices from different manufacturers, by defining communication services and protocols to ensure seamless integration. For HPPs, where a diverse set of equipment may be sourced from different manufacturers, interoperability is crucial. IEC 61850 promotes a plug-and-play environment, simplifying the integration of components from various vendors. Moreover, the IEC 61850 introduces SCL for the standardized description of substation configurations. It allows for the exchange of configuration information between devices. HPPs benefit from SCL by having a standardized way to describe the configuration of devices, making it easier to manage and maintain configurations across the plant. The IEC 61850 includes Sampled Values (SV) and Generic Object-Oriented Substation Events (GOOSE) for real-time data exchange and fast event reporting. In HPPs, SV and GOOSE enable real-time monitoring and control, thus allowing for swift responses to changing conditions. This is particularly important in scenarios where quick decision-making is critical.

In relation to cybersecurity concerns, the IEC 61850 defines security mechanisms and communication protocols to protect critical infrastructure efficiently addressing such concerns by. Given the increasing importance of cybersecurity in industrial control systems, IEC 61850 can be utilised in HPPs to implement robust security measures, safeguarding against cyber threats and ensuring the reliability of operations. Moreover, the IEC 61850 accommodates the integration of Distributed Energy Resources (DER) aligning with the growing trend of Renewable Energy Sources (RES). As part of the broader energy landscape, HPPs can leverage IEC 61850 to seamlessly integrate with other renewable sources, contributing to a more interconnected and efficient power grid. Lastly, the standardized data representation in IEC 61850 facilitates improved maintenance and diagnostics by providing a clear and consistent view of the system's state. HPPs can benefit from enhanced maintenance capabilities, allowing for more effective monitoring, diagnostics, and predictive maintenance, (being also an aspect of the iAMP-Hydro solution).

4.1.1.3 IEC 61362:2012

The IEC 61362 Ed. 2.0 b:2012, "Guide to Specification of Hydraulic Turbine Governing Systems" encompasses essential technical details needed to outline hydraulic turbine governing systems and establish their performance standards. It aims to standardize and simplify the process of selecting applicable parameters for bidding specifications and technical proposals, providing a foundation for establishing technical warranties. The focus of this standard is primarily on the turbine governing mechanisms. It also offers insights into the control loops at the plant level and primary and secondary frequency control for enhanced comprehension, without claiming comprehensiveness. Key areas addressed include: Control of speed, power, water level, gate opening, and flow for both reaction and impulse turbines, including those with double regulation; Sources of actuation energy; and Emergency shutdown safety mechanisms. To assist in the development of specifications, this standard also includes data sheets meant to be completed by both the customer and supplier throughout different phases of the project and contract. However, acceptance testing, specific testing methodologies, and warranties fall beyond this guide's purview,

4.1.1.4 IEC 62325/61970/61968

The IEC 62325, IEC 61970, and IEC 61968 are international standards related to the energy sector and power systems. These standards collectively contribute to the standardization and interoperability of systems within the energy sector. They address aspects such as communication protocols, information modeling, and application integration to enhance the efficiency and reliability of energy systems.

More specifically, the **IEC 62325-301:2018**, titled "Framework for energy market communications - Part 301: Common information model (CIM) extensions for markets" specifies the CIM for energy market communications. This framework empowers applications or systems to seamlessly access public data and engage in information exchange, regardless of the internal representation of such data. The abstract nature of the object classes within the CIM extends its applicability to a diverse range of applications, surpassing its initial usage in market management systems. In its latest version, this standard expands the scope of the CIM's utility by introducing features to support demand-side communication within a wholesale market, encompassing also provisions for demand-side resource registration and enrolment of market participating resources, along with support for the deployment and performance assessment of demand-side resources. Additionally, this last version also supports environmental (weather) data[2].

The **IEC 61970-301:2020+AMD1:2022** titled "Energy management system application program interface (EMS-API) - Part 301: Common information model (CIM) base"; defines a CIM as a standardized representation of information about the operation of the power system. This standard facilitates interoperability among different energy management systems, allowing for the exchange of information in a consistent format; by establishing a standardized method for expressing power system resources through object classes and attributes, as well as specifying their interrelationships [2].

The **IEC 61968-100:2022,** titled "Application integration at electric utilities - System interfaces for distribution management - Part 100: IEC implementation profiles for application integration"; focuses on application integration in electric utilities, including distribution management systems, energy management systems, and other related applications. It defines how messages may be exchanged between cooperating systems in order to facilitate the transfer of application-specific data [6].

The IEC standards 62325, 61970, and 61968 play significant roles in the energy sector, particularly in enhancing the standardisation and interoperability of power systems, which can directly impact projects like iAMP-Hydro. Overall, by adopting these IEC standards' notions, we can enhance its integration (e.g., achieve better integration with energy markets, allowing for efficient data exchange about hydroelectric generation), improve interoperability by supporting the integration of hydroelectric power into the overall energy management system, for real-time decision-making and optimisation; and overall achieve efficiency within the broader electric power infrastructure, contributing to more sustainable and effective water and energy resource management.

4.1.1.5 IEC 62270:2013

The IEC 62270 Ed.2.0 b:2013, "Guide for Computer-Based Control for Hydroelectric Power Plant Automation", covers the use, design, and implementation of computer-based control systems for the automation of hydroelectric plants, addressing the functional capabilities, performance standards, interfaces, hardware considerations, and training required for operators. Additionally, it provides guidelines for system testing and approval, although it does not cover the electrical protective systems for generators and step-up transformers. In its latest edition, this standard introduces significant revisions including modernized system architecture, updated communication protocols, and interface requirements.

4.1.1.6 IEC 62315:2024

The IEC 62351:2024 SER "Power systems management and associated information exchange - Data and communications security - ALL PARTS" is a set of standards dedicated to safeguarding control centers and communication networks within power systems; recommending that security measures are integrated from the very inception of a system. This implies that even though existing infrastructure may utilize older legacy systems, they can still adopt and benefit from security-by-design principles. At its core, the IEC 62351 provides comprehensive recommendations for the protection of energy management systems and the secure transmission of energy-related information. It outlines the framework for system architecture and proposes a set of effective security measures to be implemented within widely-used protocols, aiming to safeguard the confidentiality, integrity, and availability of data.

4.1.1.7 IEC 62443-2-1

The IEC 62443-2-1, titled: "Industrial communication networks - Network and system security - Part 2-1: Establishing an industrial automation and control system security program", defines the elements necessary to establish a cyber security management system (CSMS) for industrial automation and control systems (IACS) and provides guidance on how to develop those elements [7]. The objective is to ensure the security of these systems against cyber threats and vulnerabilities which are particularly critical in sectors such as hydro energy where the consequences of security breaches can be severe, including operational disruptions, safety hazards, and environmental damage.

The IEC 62443-2-1 can serve as a valuable resource for hydro energy organisations/operators seeking to enhance the cybersecurity posture of their control systems and mitigate the risks associated with cyber-attacks and operational disruptions. The use of this standard can assist in protecting critical assets such as hydroelectric dams, power generation equipment, and control systems from cyber threats; while ensuring the continuous and secure operation of the hydro energy facilities by mitigating the risks of cyber-attacks and disruptions. Moreover, it can further support existing safety and environmental protection measures, by safeguarding control systems against unauthorized access or manipulation. Moreover, this standard, offers a framework for meeting regulatory and industry standards related to cybersecurity in the energy sector. Overall, by implementing this standard's guidelines and best practices, the hydro energy sector stakeholders can strengthen the resilience and reliability of their infrastructure in the face of evolving cybersecurity threats.

In addition to the above the IEEE and IEC have developed several key standards towards to ensuring the safe and efficient operation of Hydroelectric power plants, including:

4.1.1.8 IEC 63132-1:2020

The IEC 63132-1:2020, "Guidance for installation procedures and tolerances of hydroelectric machines - Part 1: General aspects" provides a general framework defining the procedures and tolerances involved in installing hydroelectric turbines and generators. It outlines a standard assembly process, acknowledging that variations might be necessary due to factors such as machine size, design, powerhouse layout, and delivery schedules. The standard emphasises that its guidelines become obligatory only when agreed upon by both contracting parties, noting that it does not cover installations for refurbishment or small hydro projects. It also clarifies that commercial matters are generally excluded unless directly related to the installation process. Tolerances outlined are based on industry best practices and experience, with a note that they may differ from those in other standards. The document also specifies that when information, documents, or drawings are required from manufacturers, each manufacturer is responsible for providing details pertinent to their supplied components.

4.1.1.9 IEEE/IEC 63198-2775:2023

The IEC/IEEE 63198-2775:2023, titled "Technical Guidelines for Smart Hydroelectric Power Plant" outlines the integrated control and management system for smart hydroelectric power plants and plant groups which, employ state-of-the-art and widely accepted digital equipment. The provided descriptions are universally applicable to all types of hydroelectric power plants, excluding tidal and ocean power plants. This standard is based on internationally standardized communication models, and encompasses guidelines for communication networks, sensors, local monitoring and control equipment, Integrated Control and Management Platform (ICAMP), and intelligent applications. Notably, it also addresses cybersecurity considerations. The content of this document anticipates the evolving landscape of fully digitalized power plants equipped with digitalized sensors and actuators, as well as the intelligent control and management of power plants utilizing existing instrumentation[8].

4.1.1.10 IEEE 1010-2022

The IEEE 1010-2022, "IEEE Guide for Control of Hydroelectric Power Plants" serves as a reference guide for engineers in the hydroelectric field, outlining current industry standards for control system logic, configurations, and modes in hydroelectric power plant operations. It details the control and monitoring specifications for both conventional and pumped-storage hydro facilities. Furthermore, this guide outlines common practices for local and remote management, elaborating on the control interfaces for plant machinery, and specifying the prerequisites for local and remote operations control.

4.1.1.11IEEE 1020-2023

The IEEE 1020-2023, "IEEE Guide for Control of Small (100 kVA to 5 MVA) Hydroelectric Power Plants" provides detailed guidelines for the electrical control, protection, and monitoring requirements of equipment and systems associated with small hydroelectric power plants. This standard is designed to assist in ensuring the safe, efficient, and effective operation of small-scale hydroelectric installations, covering a range of aspects from system design to operational protocols.; intended for use by engineers, operators, and other professionals involved in the development and maintenance of hydroelectric power systems within the specified capacity range.

4.1.1.12IEEE 1095-2012

The IEEE 1095-2012, titled "IEEE Guide for the Installation of Vertical Generators and Generator/Motors for Hydroelectric Applications" which defines procedures for installation all types of synchronous generators and generator/motors rated 5 MVA and above to be coupled to hydraulic turbines or hydraulic pump/turbines having vertical shafts;

4.1.1.13 IEEE 1147-2021

The IEEE 1147-2021, titled "IEEE Guide for the Rehabilitation of Hydroelectric Power Plants" is designed for the hydroelectric power sector, aiming to aid owners, operators, and designers of hydroelectric plants in conducting both economic (feasibility) studies and technical evaluations (focusing on electrical components) for the rehabilitation of existing facilities. It primarily deals with traditional hydropower systems, encompassing all generation equipment up to the main transformer and standard auxiliary systems. While some sections may apply to pumped storage, specific aspects unique to that system are not addressed.

4.1.1.14IEEE 1248-2020

The IEEE 1248-2020, "IEEE Guide for the Commissioning of Electrical Systems in Hydroelectric Power Plants" which serves as a guide for plant owners, designers, and contractors involved in the commissioning of electrical systems of hydroelectric plants, providing procedures for inspections and tests for use,following the completion of the installation of components and systems through to commercial operation.

4.1.2 ISO standards

The International Organisation for Standards (ISO) provides also support to the hydropower sector including the following standards:

4.1.2.1 ISO IWA 33

The three-part ISO IWA 33 "Technical guidelines for the development of small hydropower plants" (SHP), series was developed with the input of around 80 international experts and 40 international agencies and specifies the general principles and basic requirements of design for SHP projects up to 30 Mwe. More specifically, the:

- **ISO IWA 33-1:2019** provides a technical vocabulary, defining the professional technical terms and definitions commonly used for SHP plants.
- **ISO IWA 33-2:2019**, outlines the fundamental principles and strategies for selecting sites for SHP projects; including approaches, steps, and required outcomes for identifying suitable locations for SHP facilities
- **ISO IWA 33-3:2021**, provides the technical guidelines for SHP projects and addresses various aspects crucial, such as hydrology, geology, energy calculations, project layout, hydraulics, electromechanical equipment selection, and construction planning, among others. It is designed to be site-specific, meaning its principles and requirements should be applied according to the needs of the proposed hydropower plant.

4.1.2.2 ISO 13373-7:2017

The ISO 13373-7:2017, titled "Condition Monitoring and Diagnostics of Machines - Vibration Condition Monitoring - Part 7: Diagnostic Techniques for Machine Sets in Hydraulic Power Generating and Pump-Storage Plants" provides a framework for conducting vibration diagnostics in various machinery used in hydraulic power generation and pump-storage facilities (hydropower units). This standard is addressed to those involved in condition monitoring, such as engineers and technicians, offering a detailed, vibration-based methodology for identifying faults. Additionally, the standard contains examples across different types of machinery and components, showcasing common fault indicators.

4.1.2.3 ISO 20816-5:2018

The ISO 20816-5:2018, titled "Mechanical vibration - Measurement and evaluation of machine vibration - Part 5: Machine sets in hydraulic power generating and pump-storage plants", outlines guidelines for analysing vibration data collected from bearings, bearing pedestals, or housings, and relative shaft vibration in machine sets within hydraulic power generating and pump-storage plants during standard operations

4.1.2.4 ISO 19283:2020

The ISO 19283:2020, titled "Condition monitoring and diagnostics of machines | Hydroelectric generating units" provides guidelines for condition monitoring techniques aimed at identifying and diagnosing machine faults in hydroelectric generating units, focusing on components most prone to failure. It aims to enhance the reliability of these monitoring methods and to establish common understanding and cooperation among hydropower stakeholders including end-users, contractors, and manufacturers. It specifically addresses components like generators, shaft/bearing assemblies, runners, penstocks, and certain parts of the hydraulic circuit, primarily targeting medium to large hydro units, though applicable to smaller ones. The document details also online and portable diagnostic techniques for operational units, excluding tests conducted during shutdowns, one-time tests, and systems outside the scope like transmission and civil structures.

4.1.2.5 ISO 14001:2015

The ISO 14001:2015, titled "Environmental management systems - Requirements with guidance for use" constitutes the internationally recognized standard for environmental management systems (EMS). This standard offers a structure for organisations to develop, deploy, and enhance their EMS, thereby boosting their environmental outcomes. Compliance with this standard enables organisations to adopt forward-looking strategies to reduce their environmental impact, adhere to applicable regulations, and fulfil their environmental goals. The standard covers a range of areas including the efficient use of resources, waste management, assessment of environmental performance, and engagement of stakeholders in environmental pledges. Its relevance to the iAMP Hydro project lies in providing a systematic approach to managing environmental aspects of hydropower operations, ensuring sustainable practices, and enhancing overall environmental stewardship within the project's framework.

4.1.2.6 ISO 14002-2:2023

The ISO 14002-2:2023,titled "Environmental management systems Guidelines for using ISO 14001 to address environmental aspects and conditions within an environmental topic area - Part 2: Water" addresses issues for environmental management related to water quantity and quality, such as water withdrawal, efficient use of water, and water discharge, as well as approaches to cope with water-related events such as flooding and droughts. The document considers the interconnections of water with other environmental media and takes a holistic approach to the management of water due to its impacts on ecosystems, ecosystem services, related biodiversity, as well as human life and well-being.

4.1.2.7 ISO/IEC 30179:2023

The ISO/IEC 30179:2023, titled **"**Internet of Things (IoT) **-** Overview and general requirements of IoT system for ecological environment monitoring", specifies the Internet of Things system for ecological environment monitoring in terms of the following: – system infrastructure and system entities of the IoT system for ecological environment monitoring for natural entities such as air, water, soil, living organisms; and – the general requirements of the IoT system for ecological environment monitoring.

4.1.2.8 ISO 13379-1:2017

The ISO 13379-1:2012, "Condition monitoring and diagnostics of machines - Data interpretation and diagnostics techniques - Part 1: General guidelines". This section of ISO 13379 provides universal guidelines for interpreting data and diagnosing machinery issues, aimed at standardizing concepts in machine diagnostics for users and manufacturers. It helps prepare technical specifications for further machine condition assessments and outlines methods for identifying faults in various industrial machines. Note that this standard is set to be updated by the under-development ISO/CD 13379-1.

4.1.2.9 ISO 14046:2014:

The ISO 14046:2014, titled "Environmental management - Water footprint - Principles, requirements and guidelines", specifies principles, requirements, and guidelines related to water footprint assessment of products, processes and organizations based on life cycle assessment (LCA). This standard aims to assess the environmental impacts related to water use and provide a framework for water footprint assessment that includes setting the scope, conducting a water usage inventory, assessing potential impacts, and interpreting the results. This standard helps organisations understand and reduce their water usage, contributing to more sustainable water management practices. It is applicable to any organisation looking to evaluate and improve its water footprint, regardless of size, location, or industry.

4.1.2.10 ISO 17359:2018

The ISO 17359:2018,titled "Condition monitoring and diagnostics of machines - General guidelines" It outlines a framework for establishing, implementing, evaluating, and improving a condition monitoring program for various types of machinery. This includes setting up the program's scope, planning the monitoring activities, collecting and analysing data, and making maintenance decisions based on the findings.

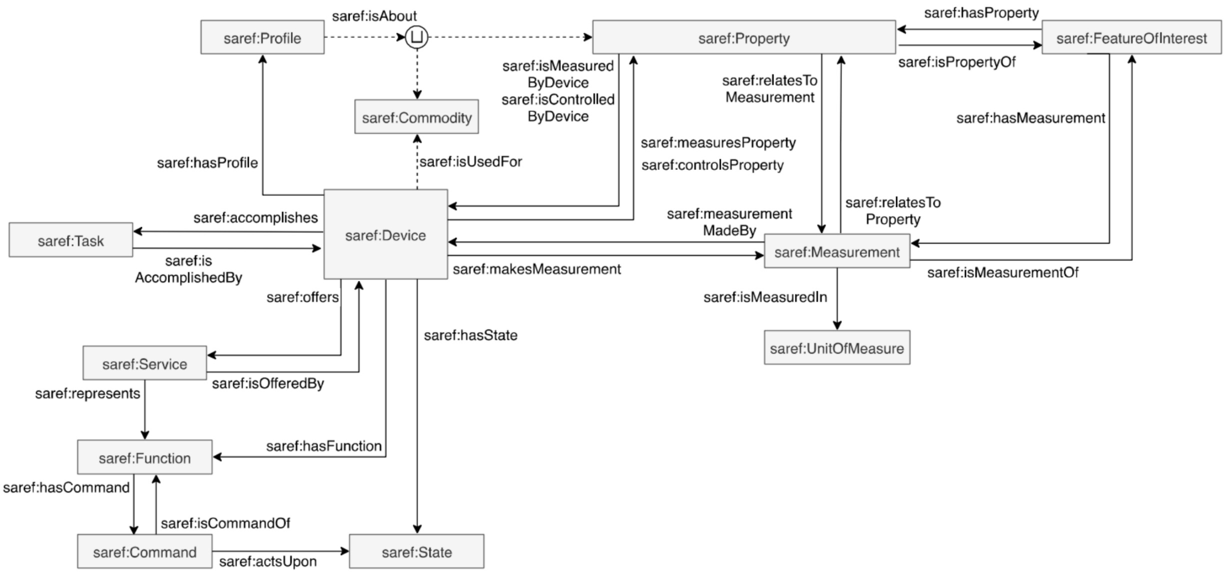
In the context of HPPs and hydro power generation, the ISO 17359:2018 is considered as highly applicable as it can help in ensuring the reliability and efficiency of critical components such as turbines, generators, and auxiliary systems. By adopting the guidelines set out in this standard, HPPs can effectively monitor the condition of their machinery, predict potential failures, and perform maintenance proactively. This not only helps in minimizing downtime and maintaining continuous power generation but can also extend the lifespan of the equipment and improve safety and environmental compliance.

4.2 Review of Existing Ontologies:

4.2.1 SAREF

The Smart Applications REFerence (SAREF)[1] ontology is a family of ontologies developed by the Dutch institute TNO and published by the European Telecommunications Standards Institute (ETSI), to assist in utilising diverse assets within the domain of smart applications, Figure 2. SAREF aims to offer modular and adaptable components of the ontology tailored to user requirements and is constructed following several fundamental guidelines. Primarily, it allows for the reuse and realignment of concepts from existing assets, while it permits the segregation and reassembly of its various elements to cater to specific user needs. Additionally, it supports extensions, ensuring that the update, identification, and modification of any errors remain a straightforward and easily manageable processes. This ontological framework facilitates the development of abstraction layers for devices and technologies, alongside their respective universal Application Programming Interfaces (APIs), all without requiring a detailed understanding of particular standards.

Figure 2. Overview of the SAREF ontology[1]



The SAREF ontology dictates a single mapping set for each asset, eliminating the need for unique mappings between each asset pair. While assets may have overlapping fundamental concepts, they typically employ distinct terminologies and data models, as such SAREF enables various assets to maintain their individual terminologies and models yet establishes connections among them via the use of their common semantics.

The adoption of the SAREF in the domain of hydropower plant operations (and in the broader scope of iAMP-Hydro) can significantly enhance operational efficiency, safety, and interoperability, both internally and with external systems. By adopting a SAREF, hydropower plants can improve data management, streamline communications between devices, and facilitate more informed decision-making, leading to more sustainable and efficient operations. SAREF can be applied towards facilitating the integration of diverse HPPs devices, sensors, actuators, etc. by providing a common language and set of concepts. For example, devices like sensors in the plant can be classified under SAREF's 'Device' category, and their operational commands can be standardized (e.g., StartCommand, StopCommand for turbines). In addition, adoption of SAREF can facilitate data harmonisation by offering a unified model that represents core concepts (e.g., Energy, Pressure, etc.). This uniformity can enable more effective data sharing and analysis across the plant's ecosystem; while also operational efficiency (e.g., in terms of optimised water flow and energy production) can also be since by applying SAREF,

[1] https://saref.etsi.org

operations can be streamlined by providing clear and consistent definitions for the properties and actions of each device. In addition, and relevant to the scope of the iAMP-Hydro project, utilizing SAREF's predictive maintenance models can be developed using the standardized data collected from various sensors and devices. By analysing this data, plant operators can predict potential failures or safety issues before they occur, thus reducing downtime and enhancing safety protocols.

4.2.2 SAREF4ENER

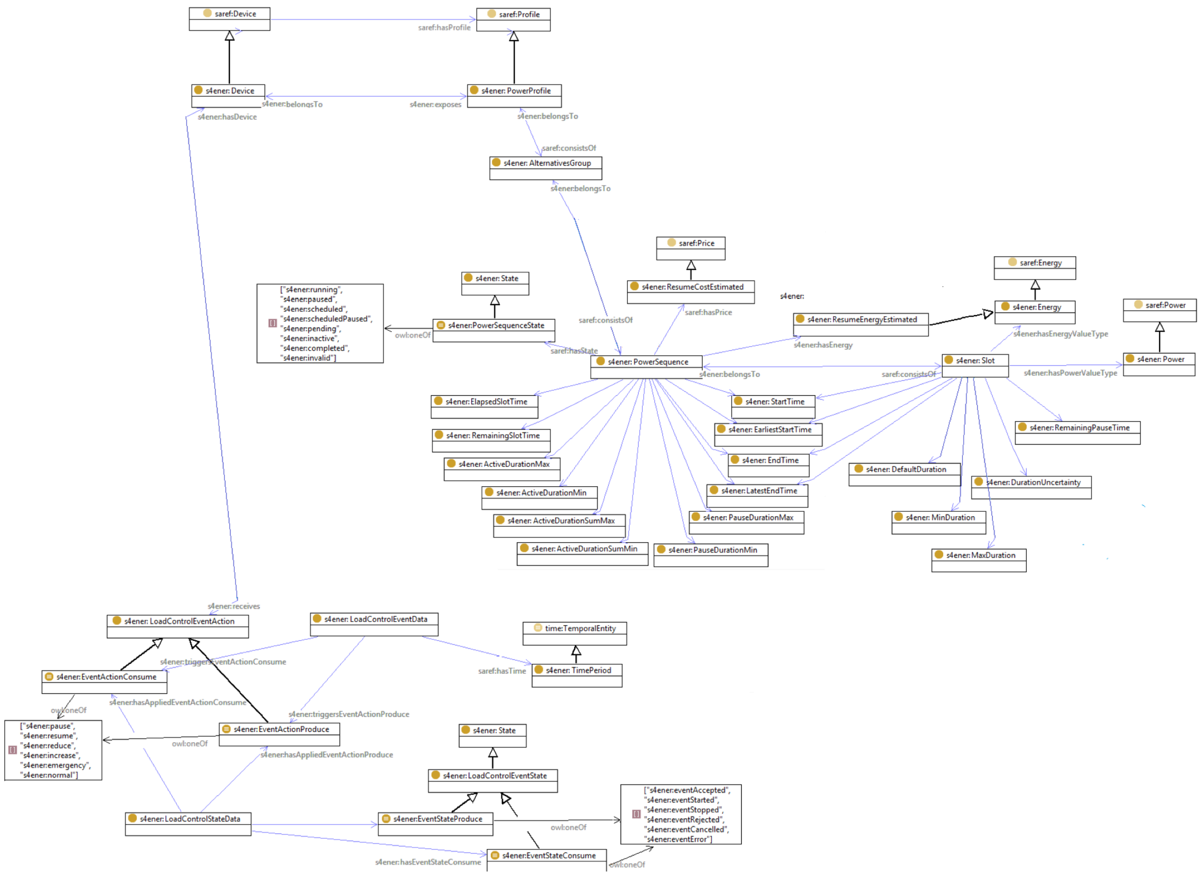
SAREF for the energy system (SARF4ENER)[2], is a specific extension of SAREF focusing on energy-related applications and the energy domain in general. It was originally developed through the collaborative efforts of two prominent organizations, Energy@Home and EEBus, which aimed at facilitating interoperability between their distinct data models. In general, this ontology provides a standardized and semantically rich framework for representing energy-related information, making it a valuable tool for achieving seamless communication and data interoperability in the energy domain. The SAREF4ENER ontology extension, as illustrated in Figure 3, introduces 63 new classes, 17 object properties, and 40 data type properties, with a focus on demand response scenarios. In such scenarios, consumers have the opportunity to contribute to the Smart Grid's optimization of energy use / production by changing their demand patterns during peak times, thus appropriately managing their smart home devices.

SAREF4ENER presents a promising solution for addressing the complexities of hydroelectric energy systems. The application of SAREF4ENER in HPPs can enhance data exchange, device interoperability, and overall system efficiency, contributing to the sustainable and optimised operation of hydroelectric power generation. SAREF4ENER can be applied to represent hydrological data within HPPs, as it includes the modeling of water levels, river flow rates, and precipitation data, providing a standardized format for communication between hydrological monitoring systems. Through the provision of detailed device descriptions, SAREF4ENER facilitates interoperability among various devices used in HPPs, such as sensors for water level measurement, turbines, and control systems. This ensures that devices from different manufacturers can communicate seamlessly. Moreover, SAREF4ENER supports the modeling of energy generation patterns over time. As such, SAREF4ENER can be utilised in HPPs' operations to represent the dynamic nature of energy production based on variations in water availability and turbine operations.

Overall, the SAREF4ENER's compatibility with broader SAREF extensions can enable also HPPs to integrate with smart grids efficiently, which is vital for optimising energy distribution and ensuring the coordinated operation of hydroelectric systems within larger energy networks.

[2] https://saref.etsi.org/saref4ener/v1.1.2/

Figure 3. Overview of SARE4ENER[3]



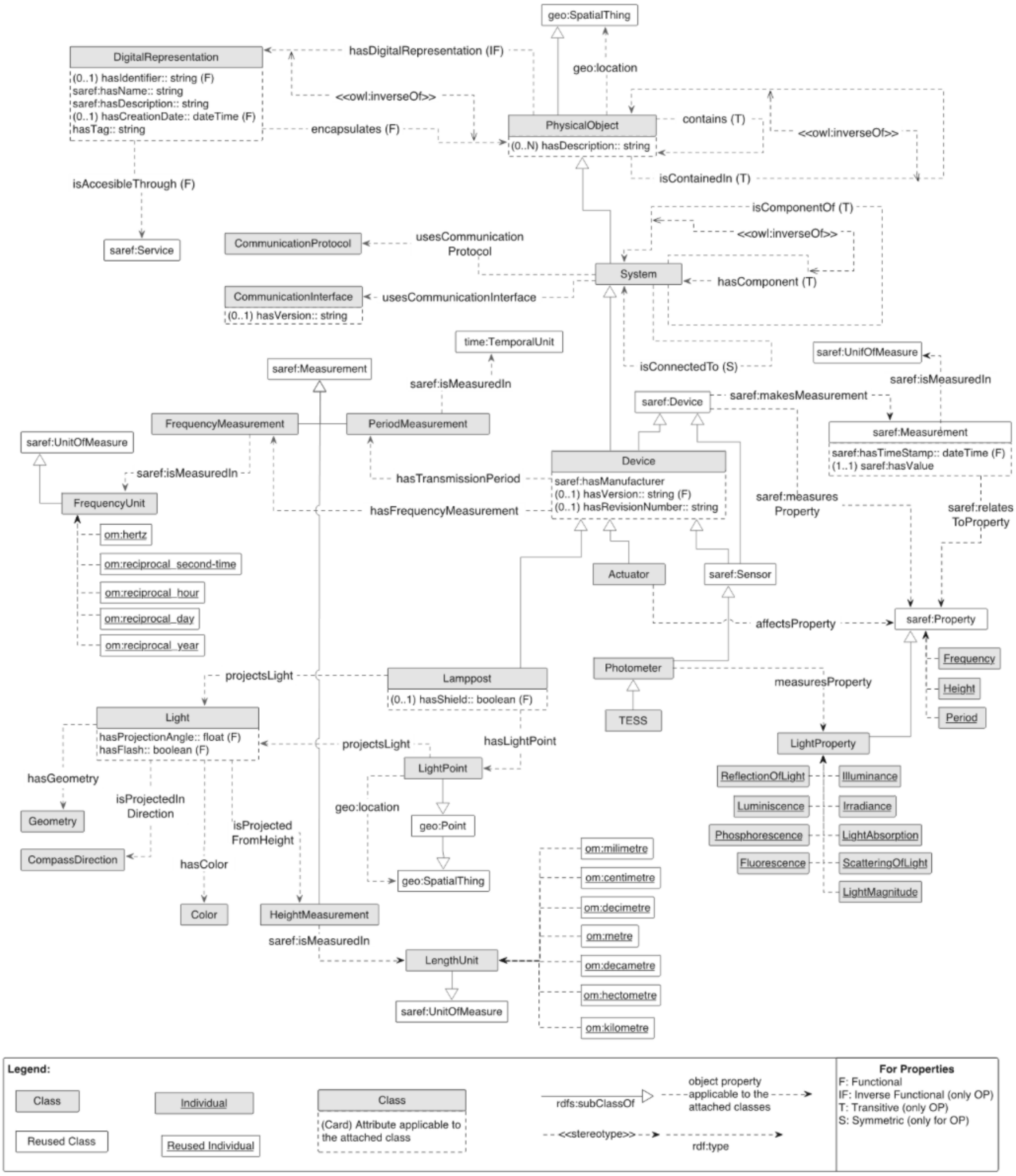
4.2.3 SAREF4ENVI

SAREF4ENVI represents another specific extension of the SAREF ontology specifically tailored to address the domain of environmental monitoring and management. It extends the core concepts of SAREF to cover aspects related to environmental sensors, data collection, analysis, and decision-making processes in environmental applications[4]. SAREF4ENVI is an OWL-DL ontology that extends SAREF with 32 classes, 24 object properties, 13 data type properties, and 24 individuals. A high-level view of the SAREF4ENVI ontology extension is presented in Figure 4.

Overall, SAREF4ENVI focuses on environmental monitoring and management, by providing a standardized framework for describing environmental monitoring and management systems. This includes sensors and devices used to collect environmental data, such as temperature, humidity, air quality, pollution levels, and weather conditions. Moreover, it defines semantic models and vocabularies to represent the functionalities, properties, and relationships of environmental monitoring devices and systems in a machine-readable format. This enables interoperability and seamless integration of diverse environmental monitoring solutions from different manufacturers and vendors

[3] https://saref.etsi.org/saref4ener/v1.1.2/
[4] https://saref.etsi.org/extensions.html#SAREF4ENVI

Figure 4. Overview of SAREF4ENVI[5]



By adopting SAREF4ENVI, stakeholders in the hydro power domain can ensure that data collected from various sensors and devices are represented using a common semantic model. This facilitates data integration, aggregation, and exchange between different systems and applications, enabling more efficient environmental monitoring and decision-making processes. Overall, adoption of the SAREF4ENVI ontology extension can be instrumental in managing the various assets of a hydropower plant, such as turbines, generators, and gates. It can enhance environmental monitoring, compliance, and management and supports the integration of environmental considerations into all aspects of hydropower operations, from water management to asset maintenance and energy production, leading to more sustainable and efficient operations.

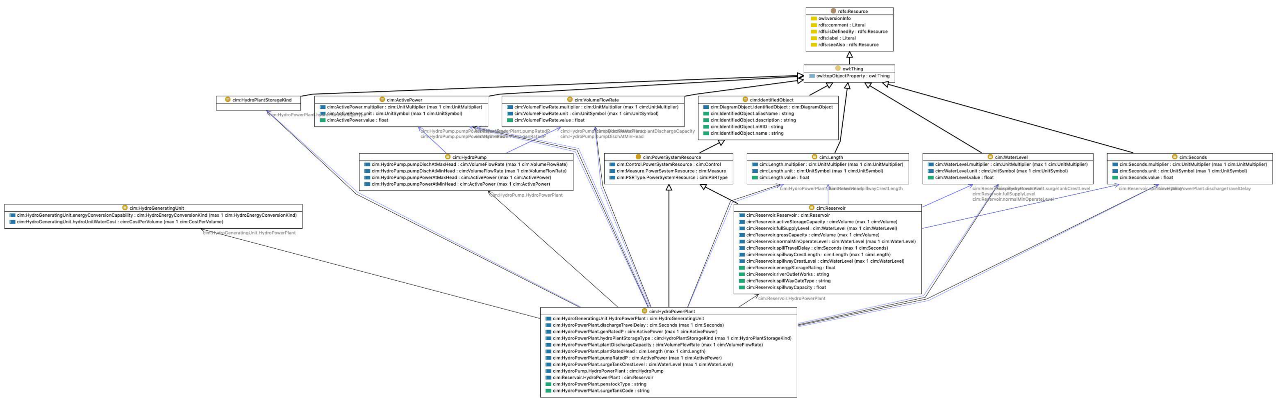
[5] https://saref.etsi.org/saref4envi/v1.1.2/

4.2.4 CIM (IEC 61970 part 3)

The IEC 61970-301:2020+AMD1:2022 "CSV Energy management system application program interface (EMS-API) - Part 301 Common information model (CIM) base [5], and the IEC 61970-302:2024 Part 302: "Common information model (CIM) dynamics [6] refers to the Common Information Model developed by the IEC's Technical Committee 57 (TC57). The CIM provides a standardized framework for describing the electrical network, enabling efficient data exchange and integration between different utility systems, particularly in the context of electrical utilities and energy markets. This ontology covers various aspects of electrical systems, from generation and transmission to distribution and market communication. It defines common terms and relationships used in the industry, ensuring that different systems and organizations can understand and use the data consistently. Indeed

Figure 5 illustrates the CIM, which encompasses the 'HydroPowerPlant' class as a critical component within its framework. This depiction is part of the broader ontological structure, methodically detailing the various components and relationships inherent to hydroelectric power generation facilities.

Figure 5. CIM: Ontological Structure of Hydro Power Plant and components[6]



The CIM is important for HPPs, offering a standardized approach for data management and communication, which aids in grid integration, operational efficiency, and compliance with regulatory standards. More specifically, by adopting the CIM, HPPs can improve their operational efficiency through better integration with the electrical grid and more effective data exchange. The CIM ontology supports also the integration of hydro power as a renewable energy source into the broader energy market, aiding in balancing supply and demand. Facilitates the participation of HPPs in grid stability and reliability programs, thanks to standardized communication and data models. Moreover, it can assist hydro power operators comply with industry regulations and standards by providing a framework for consistent data management and reporting.

4.2.5 WHOS Hydrological ontology

The WHOS Hydrological Ontology[7] comprises of concepts relating to the hydrology domain, their properties and their interconnections. It includes concepts from the Ontology of the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) including their attributes and additional synonyms in multiple languages, and their relationships among them.

Overall, it serves to enhance the exchange of hydrological data and promote semantic interoperability across various scales, from local to global, thus it is a dynamic resource that is continuously enhanced and updated to reflect advancements in the field. Towards enabling machine-to-machine interaction, the WHOS Hydrological ontology is available through the semantic query language - SPARQL, where users can query it. Figure 6 presents a generic view of the WHO's Hydrological Ontology. In this

[6] https://ontology.tno.nl/IEC_CIM/cim_HydroPowerPlant.html
[7] https://community.wmo.int/en/whos-hydrological-ontology

representation, concepts are illustrated as circles, and the relationships between them are depicted as arrows.

Figure 6. Generic view of WHOS Hydrological Ontology, through an RDF viewer tool[8]

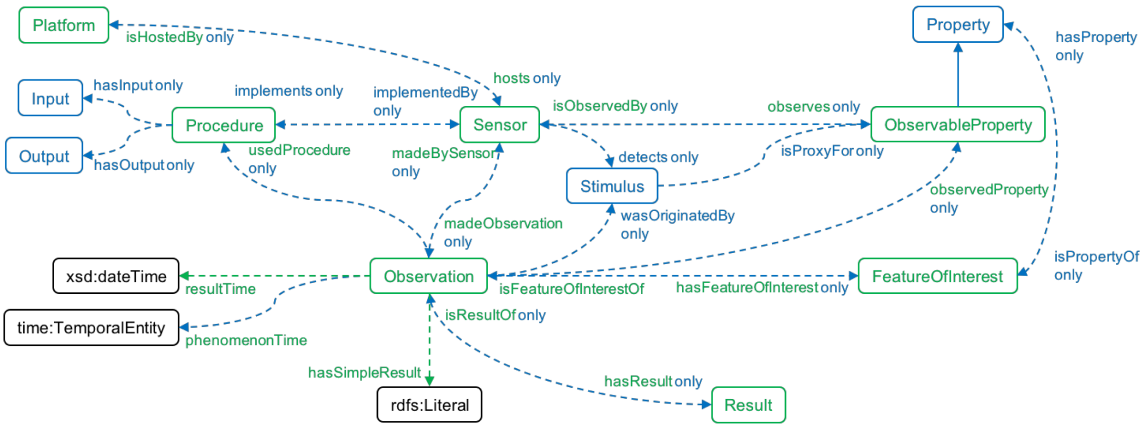
4.2.6 SSN-SOSA ontology

The Semantic Sensor Network (SSN) ontology serves as an additional framework for defining sensors and their observations/measurements, encompassing associated procedures, the targeted features of interest, utilised samples, and the observed properties, alongside actuators. Although similar to SAREF and both built on semantic web technologies, utilising RDF (Resource Description Framework) for data representation; SSN can be considered as more specific to environmental monitoring, including hydrology, weather, and other natural phenomena. It focuses on sensors, observations, samples, and actuators, making it well-suited for scientific and environmental application.

SSN adopts a modular architecture, both horizontally and vertically, incorporating a streamlined yet comprehensive core ontology known as SOSA (Sensor, Observation, Sample, and Actuator) to address fundamental classes and properties, Figure 7. By encompassing diverse scopes and levels of axiomatization, SSN and SOSA facilitate a broad spectrum of applications and scenarios. These include but are not limited to satellite imagery, extensive scientific monitoring endeavours, industrial and residential infrastructures, social sensing initiatives, citizen science projects, ontology engineering driven by observation, and the interconnectedness of the Web of Things.

[8] http://hydro.geodab.eu/ontology-browser/hydro-ontology.html?http://hydro.geodab.eu/hydro-ontology/concept/scheme

Figure 7. Semantic Sensor Network: Observation Model[9]

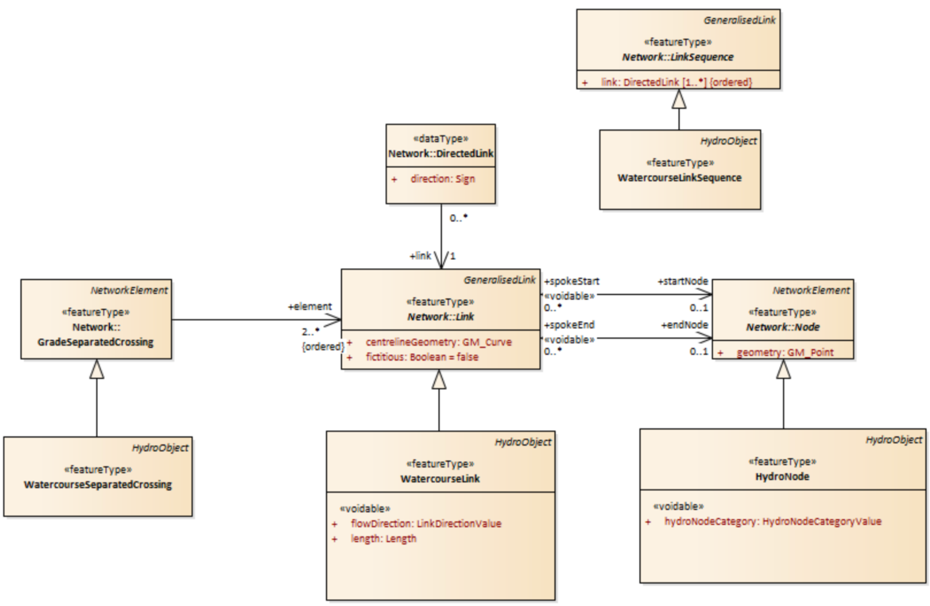


4.3 Review of Existing Data Models/Schemas

4.3.1 INSPIRE Hydro Network Model

The objective of the Infrastructure for Spatial Information in Europe (INSPIRE) Directive is to establish a European Union Spatial Data Infrastructure (SDI), serving the needs of EU environmental policies and any initiatives that might influence the environment. The goal is to foster the exchange of environmental spatial data among public sector entities, enhance public accessibility to spatial information throughout Europe, and aid in cross-border policymaking. INSPIRE builds upon the spatial information infrastructures already implemented and managed by the member states of the EU. As part of the INSPIRE Directive, the INSPIRE data specifications Technical Guidelines define GML application schemas (xml schemas) as the default encoding for all INSPIRE spatial data themes, Figure 8. These xml schemas are available in the INSPIRE schema repository[10].

Figure 8. INSPIRE Hydro - Network: spatial object types: Class diagram [11]



More specific, to the objectives of the iAMP-Hydro, INSPIRE includes the Hydro-Network model, a data model for the representation of hydrographic networks, towards facilitating the exchange of hydrological data among EU member states. The Hydro-Network schema contains spatial objects associated with a connected hydrographic network representation of waterways, lakes etc. A glimpse

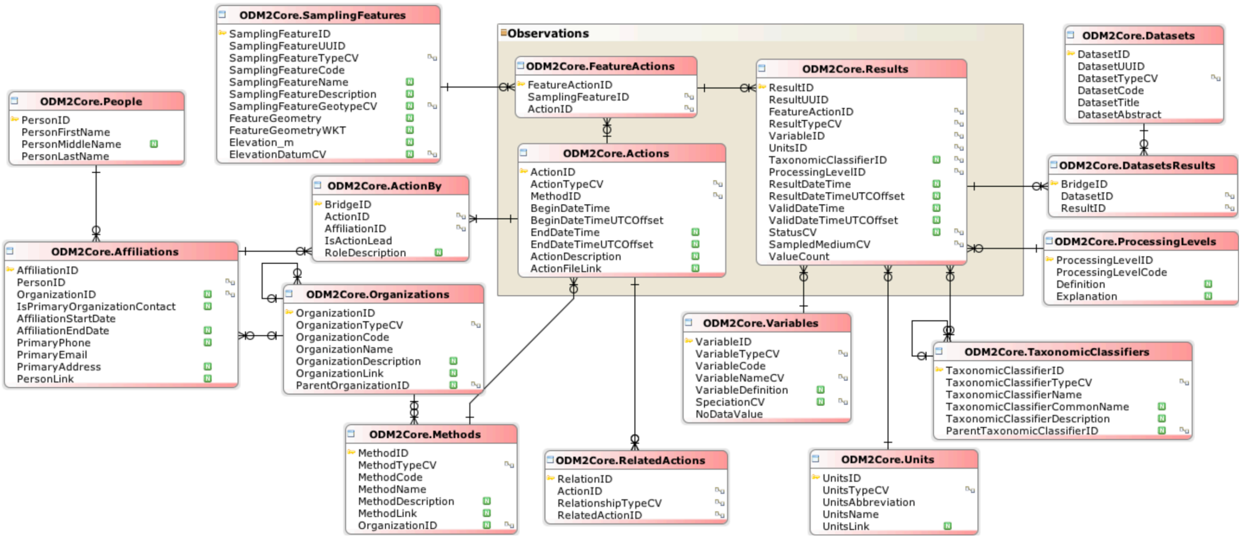
[9] https://www.w3.org/TR/vocab-ssn/images/SSN-Observation.png
[10] https://knowledge-base.inspire.ec.europa.eu/tools/xml-schemas_en
[11] https://inspire-mif.github.io/uml-models/approved/html/index.htm?guid=CAB1DAF6-2F62-4151-8C10-2B34A62A0431

of the INSPIRE Hydro-Network can be seen in the above figure, presenting the class diagram for key spatial hydrological objects.

4.3.2 ODM2

The Observations Data Model 2[12] (ODM2) is both an information model for spatially discrete, feature-based earth observations that is integrative and extensible and aimed at data interoperability, and a supporting software ecosystem. It incorporates principles from ODM1 and various pre-existing cyberinfrastructures to broaden its capability in consistently characterising, storing, handling, and encoding observational datasets for archival and transmission. ODM2 is organized with a "core" schema and multiple "extension" schemas that extend the functionality of the core. In contrast to alternative platforms, it supports a broader spectrum of observational data originating from both sensors and specimens[12]. ODM2 metadata structure provides a powerful new way to describe sampling, sensing, and analysis workflows, incorporating the following core entities: Actions; Datasets; Methods; Organizations; People; Processing Levels; Related Actions; Results; Sampling Features; Taxonomic Classifiers; Units; Variables. An overview of the ODM2 Core Schema is illustrated in Figure 9.

Figure 9. Overview of the ODM2 Core[13]



The relevance and applicability of ODM2 to the iAMP-Hydro primarily revolves around its capabilities for managing and organizing environmental data. ODM2 is designed to support the storage and retrieval of observational data, which is crucial in hydrological analysis and the management of HPPs. ODM2 can handle diverse types of data including physical, chemical, biological, and geological variables; also supporting time-series data, allowing for effective tracking and analysis of changes over time, which is essential for operational efficiency and environmental compliance. As such the data organized within ODM2 can be used to feed hydrological models and simulations, which are essential for the planning and efficient operation of HPPs.

4.3.3 OGC WaterML 2

The OGC WaterML 2 is a standardized XML-based language developed by the Open Geospatial Consortium (OGC) focused on the representation and exchange of groundwater data. It is specifically designed to facilitate the management, sharing, and integration of groundwater information across different platforms and organisations; by supporting the interchange of hydrological and hydrogeological information such as time series, hydrological features, and related metadata.

[12] https://www.odm2.org
[13] https://odm2.github.io/ODM2/schemas/ODM2_Current/diagrams/ODM2Core.html

WaterML 2.0 is part of the Hydrology Domain Working Group's efforts to improve the interoperability within the water data community.

WaterML2.0 is implemented as an application schema of the Geography Markup Language (v3.2.1), making use of the OGC Observations & Measurements standards. It is made to be a flexible schema tailored for encoding data applicable in various data sharing contexts. Examples for its use include hydrological monitoring programs, infrastructure management like dams and supply systems, cross-border data sharing, public data dissemination, disaster management improvement through information sharing, and national reporting support. The essential element of this framework is the accurate and detailed representation of time series data. Understanding these time series is essential and depends on the processes that produced them. WaterML2.0 sets the groundwork for the exchange of time series with necessary metadata, ensuring accurate machine interpretation and effective subsequent analysis. It aims to serve as a theoretical 'bridge' for existing systems, enabling them to interface with different schemas or systems while preserving data consistency.[14]

We consider the OGC WaterML 2.0 relevant to the iAMP-Hydro, as it standardizes the sharing and interpretation of water-related data across different entities. This standardization not only enhances the ability to manage time-sensitive data, e.g., water levels and flow rates, but also improves the accuracy of hydrological models, leading to more informed water management decisions. Additionally, it supports the integration of varied data sources, ensuring a comprehensive view of water resources, and facilitates real-time data exchange. Overall, GWML2 can be considered as an important standardized schema for the effective management of groundwater resources, directly impacting the efficiency, sustainability, and regulatory compliance of hydro power operations.

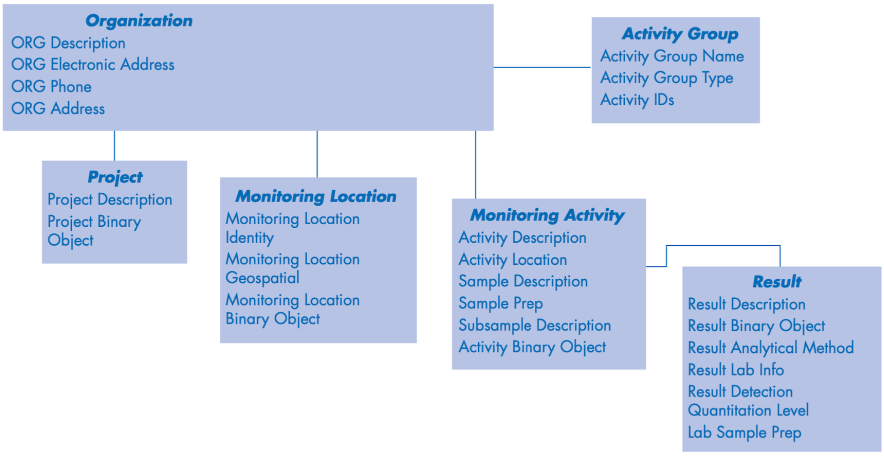
4.3.4 WQX (Water Quality Exchange)

The Water Quality Exchange (WQX) is a framework and format developed by the Environmental Protection Agency (EPA) for the purpose of facilitating the sharing and exchange of water quality data. WQX is part of the broader Water Quality Data (WQD) framework, which aligns with the National Water Quality Monitoring Council (NWQMC) standards and guidelines. It uses the standards of the National Environmental Information Exchange Network to facilitate data sharing with EPA. In essence, it serves as a mechanism for submitting and sharing water monitoring data over the internet, allowing for effective data management and interoperability between different data systems, promoting transparency and collaboration in water quality monitoring and management [11]. The WQX schema is a standard set of data formats that specify the data elements and data structure required for submission of data to EPA, Figure 10.

By using WQX, organizations can ensure that their water quality data is standardized, comparable, and easily shared with regulators, researchers, and the public.

[14] https://schemas.opengis.net/waterml/ReadMe.txt

Figure 10. Simplified graphic of the WQX schema showing core data elements



Overall, the WQX schema is designed to improve the accessibility and integration of water quality information, which is crucial for assessing the health of water bodies and for making informed environmental and public health decisions. As also shown in the above figure, the WQX is structured as follows:

- Header Information, including metadata such as the organization providing the data, the date of submission, and the geographical scope (e.g., national, state, local).
- Monitoring location information: Identifying the specific locations where water samples are collected, including details like site name, geographical coordinates (latitude and longitude), and the type of water body (e.g., river, lake, stream).
- Sample Information including details about each water sample collected, including the date and time of collection, sample type (e.g., surface water, groundwater), and sampling method.
- Results, which includes the actual data obtained from analysing the water samples. This includes the type of parameters tested (e.g., temperature, pH, dissolved oxygen, contaminants), measurement values, units of measurement, and the analysis methods undertaken.
- Quality Assurance/Quality Control (QA/QC) Information , including QA/QC procedures and results to ensure the reliability and accuracy of the data.
- Project Information, providing a description of the project or program under which the data were collected, including objectives, funding sources, and relevant documents.
- Biological Data information, applicable to projects involving biological monitoring, including details about species observed, their abundance, and the habitat conditions.

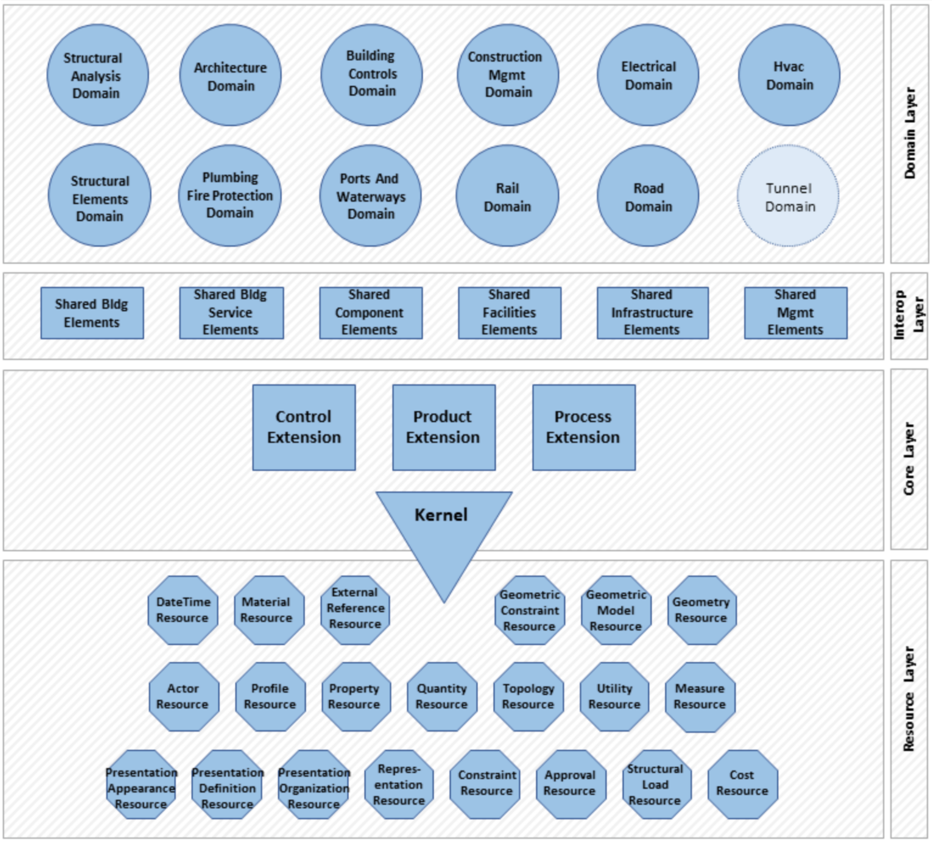
4.3.5 IFC 4.3.x

The Industry Foundation Classes (IFC) released by buildingSMART International, serve as a public standard for the exchange and sharing of Building Information Modeling (BIM) data among stakeholders involved in building projects. In its current release IFC 4.3 encompasses the data structure and reference data, represented through HTML, EXPRESS, XSD/XML, and OWL documentation and formats. In more detail, the IFC framework outlines terminology, concepts, and data items that derive from various professional, trade, and disciplinary uses within the building industry [10]. The framework is structured into four conceptual levels: the resource, core, interoperability, and domain layers, as depicted in Figure 11, where the:

- Resource layer aggregates all individual schemas detailing resource definitions.
- Core layer is made up of the primary schema and its extensions, housing the broadest entity definitions. It is important to mention that entities at the Core layer and above (this includes the Interoperability and Domain layers) are assigned a unique global identifier, with optional owner and historical data.

- Interoperability layer comprises schemas for entity definitions that are generally applicable across multiple disciplines, aiding in the cross-disciplinary exchange and sharing of construction information.
- Domain layer holds specialized entity definitions related to specific disciplines within the building sector, facilitating their exchange or sharing throughout the entire data schema.

Figure 11. Architectural Overview of IFC v4.3.x[15]



[15] https://standards.buildingsmart.org/IFC/RELEASE/IFC4_3/HTML/content/introduction.htm

5 Key Findings and Gap identification

5.1 Gap Identification

Initially, it should be emphasized that the gap analysis presented herein is strictly based on the examination of standards specific to the hydropower-energy sector and which provide semantic ontologies and data models; thus we omit from our analysis the IEEE,IEC and ISO standards outlined in Chapter 4 that are fundamentally intended as guiding frameworks and standardised methods applicable at HPPs operation, rather than including any semantic or data management aspect relevant to the scope of the iDML.

Under this context, following our review of these ontologies and data models, several gaps have been identified which may impact the efficiency and completeness of the iAMP-Hydro project, Table 6:

- Specificity to Hydropower operations: Current ontologies like SAREF, SAREF4ENER, and SAREF4ENVI, while extensive, do not fully address the specific operational, maintenance, and safety needs inherent to hydropower plant systems. There's a need for a more "hydro-specific" ontology that covers unique aspects such as turbine efficiency, reservoir management, and environmental impact. Moreover, harmonization with the IEC 61850/ IEC 61970 (CIM) standards is needed so that we can comprehensively address hydropower operations and semantically integrate them with energy and environmental concepts.
- Comprehensive Environmental Data Integration: While SAREF4ENVI and the WHOS Hydrological Ontology provide a good foundation for environmental monitoring, there's a lack in capturing the full range of environmental impacts specific to hydropower, such as aquatic biodiversity, sediment transport, and water temperature changes.
- *Real-time data Handling and predictive analytics:* Existing data models like ODM2 and WaterML 2.0 focus on data storage and sharing but lack comprehensive support for real-time data processing and predictive analytics necessary for operational efficiency and prevention of damages in hydropower plant equipment; also, very relevant to the project's tasks pertaining to predictive maintenance analytics and water flow prediction. While the SSN-SOSA ontology seems a fit here as it covers a wide range of sensors and actuators in general terms; it lacks specific concepts or fields tailored to the unique types of sensors and actuators used in hydropower plants, such as flow meters, pressure transducers, or turbine efficiency sensors.
- *Interoperability among diverse data models*: While the IEC 61970 (CIM) ontology is comprehensive regarding electrical systems, advocating interoperability through its standardized data representation, there is no specific adaptations for the unique aspects of hydropower operations, as it does not include crucial hydro-specific components like turbines, gates, and spillways, which are crucial for effective hydropower management. Moreover, there is a noticeable gap in interoperability between other reviewed data models and ontologies (e.g., between ODM2, WQX WaterML 2.0, and SAREF extensions); lacking a harmonised approach that can enables seamless data exchanges and integration across these models. Here we also consider the IFC standard relevant to the infrastructure and hydropower projects. Although only few of its concepts are relevant for iAMP-Hydro, the iDML shall also consult the IFC to ensure alignment.
- *Data security and privacy concerns*: None of the reviewed ontologies and data models specifically address the critical aspects of data security and privacy, which are paramount due to the sensitive nature of hydrological and operational data in hydropower projects.

Table 6. Gap identification of ontologies and data models analysed for iAMP-Hydro

Ontology/ Data model	Application Domain	Weaknesses	Pertinence to iAMP-Hydro
SAREF	IoT, Smart Applications	SAREF is an ontological framework strongly endorsed by the EC. It serves as an abstract layer above various communication protocols,	Although not directly transferable, its data semantic focus may aid in iAMP-Hydro framework's

		aiming to outline the fundamental concepts, their interrelations, and connections to different external frameworks specifically in the smart appliances sector. While it holds significant relevance, it is not directly transferable to iAMP-Hydro.	interoperability considerations.
SAREF4ENER	Energy	Tailored for the energy sector, may not fully address hydrological data needs or concepts specific to water management.	Relevance might be indirect, focusing primarily on energy-related aspects within HPPs contexts.
SAREF4ENVI	Environment	Focused on environmental aspects but may lack detailed water-related features and specific hydrological functionalities.	Could provide environmental context beneficial to iAMP-Hydro but may require adaptation for full relevance.
CIM (IEC 61970-301:2020+AMD1:2022 & IEC 61970-302:2024)	Energy, Utilities	Primarily designed for electrical grids and utilities management, not for water or hydrological systems.	Not directly relevant due to its primary focus on energy sectors rather than HPPs operations. Nevertheless, harmonization with this standard is important to address hydropower operations and semantically integrate them with energy and environmental concepts.
WHOS Hydrological	Hydrology	Dependence on the willingness and ability of countries and organizations to share data, which can lead to gaps in global data coverage. Moreover, potential issues with data quality, accuracy, and resolution given the broad range of data contributors.	Relevant due to its specialised focus on hydrology.
SSN-SOSA	Sensing, Observation	Generalised for sensing and observation measurements; may require customisation to fit specific hydrological monitoring requirements.	Its generic sensing framework could be adapted to support iAMP-Hydro's data collection and observational needs.
INSPIRE Hydro	Hydro Network Modelling	While aiming for interoperability, different interpretations across countries can lead to inconsistencies, affecting the seamless exchange and use of hydrological data. Also The model's strict encoding rules could also restrict flexibility in data representation and management.	Directly relevant as it specifically addresses hydro network modeling, aligning well with iAMP-Hydro requirements.

ODM2	Observational Data	Although ODM2 is designed to be extensible, its core is tailored for spatially discrete, feature-based earth observations, which might limit its applicability outside of geosciences or earth observation data.	Considered as applicable due to its focus on observational data, relevant to iAMP-Hydro's monitoring activities.
OGC WaterML 2	Water Data	WaterML 2.0 is acknowledged to be complex, making it challenging to understand, especially if detailed elements are utilised.	Applicable, as it is tailored for water data representation, relevant to the scope of the iAMP-Hydro.
WQX	Water Quality	Implementing and using WQX may require substantial technical expertise, particularly for users looking to set up custom XML submissions through Exchange Network Nodes or Node Clients. This could limit accessibility for organizations with less technical capacity.	Specifically relevant for water quality aspects within iAMP-Hydro.
IFC 4.3	Construction, Buildings	Primary designed for the AEC industry and not tailored for hydrological or environmental data.	Not directly pertinent, due to its focus on construction and infrastructure elements rather than hydrological data. Nevertheless, should HPPs building data be needed, IFC can provide a structured data modeling framework to be based on.

5.2 Main takeaways for the iDML

In addition to the aforementioned identified gaps and the insights provided in Chapter 3, we consider the following as key considerations to be addressed during the development of the iDML within the iAMP-Hydro project:

- The iDML should focus on enhancing the standardization and interoperability among different hydropower systems and equipment. Ontologies such as the one provided from IEC 61970 part 3 and the different SAREF extensions can aid in achieving unified communication protocols and the required data exchange formats; thus, it is critical to leverage their relevant information entities, their concepts along with their relationships and further enhancing them with entities, concepts that address the unique aspects of iAMP-Hydro project.
- The iDML, shall take into consideration cybersecurity best practices from standards like IEC 62443-2-1 within iDML can bolster the security posture of hydropower infrastructure against emerging threats and vulnerabilities.
- The iDML should integrate sustainability metrics and consider the guidelines from ISO 14001 and ISO 14046 to support environmentally responsible and sustainable hydropower operations.
- The iDML, shall be developed as a custom data model addressing the specificities of the iAMP-Hydro project, while following (where feasible) the structure of relevant data models and

ontologies, this can significantly improve data management, real-time monitoring, and predictive analytics.

- The iDML should be designed to be adaptable and extendable, allowing for the incorporation of future models, ontologies and practices as they emerge and evolve within the hydropower sector.

6 Conclusions

Overall, D8.1 outlines the key requirements for the data model to be utilised within the iDML for harmonising and enabling interoperability among the different data that will be leveraged within the context of the iAMP. As such D8.1 begins with presenting the methodology followed towards determining the semantic data needs of the iAMP-Hydro project and identifies preliminary High level Information entities and domain concepts that the data model should possess, as extracted from the project's objectives and from an initial data landscaping exercise undertaken.

Following this, an examination of current data models, ontologies, and standards, is conducted to identify relevant ones to the hydro-energy sector. This analysis assesses the extent to which the existing standards, data model and ontologies fulfil the unique prerequisites of the iAMP-Hydro project. It is observed that these standards, data models, and ontologies vary significantly in their scope and suitability for different needs, ranging from operational guidelines for HPPs, communication protocols to complete ontologies and data models.

Upon evaluating these elements and identifying initial shortcomings (in respect to iAMP-Hydro needs) in the existing models and ontologies' coverage, it becomes evident that no singular data model can satisfy all criteria for the iDML; therefore, the iDML is proposed to integrate the context of various relevant data models and ontologies, rather than conforming to an individual existing one. Moreover, this deliverable defines key operational requirements for the iDML, including:

- Ensuring compatibility with established systems, aiming for interoperability without replacing them, thereby maintaining or enhancing semantic data integration without compromising efficiency at a big data level.
- Maintaining clarity and precision within the iDML to avoid misinterpretations.
- Incorporating additional modeling attribute that, while may seem secondary, are essential for accurate alignment with external data sources.
- Ensuring the iDML's design incorporates extensibility to facilitate future growth and adaptations, enabling seamless extension and refinement of its capabilities.
- Implementing a strategy for regular refinement to address and rectify any emerging ambiguities due to the model's evolution.

Finally, D8.1 concludes by emphasising the necessity for the iDML to:

- Provide comprehensive coverage across all pertinent data domains to the hydro-energy industry, while meeting the specific iAMP-Hydro data modelling needs;
- Ensure effectiveness and adaptability for large-scale data environments;
- Allow for future modifications and enhancements.

References

[1] iAMP-Hydro Description of Action 2023

[2] IEC61850:2023 SER ; Communication networks and systems for power utility automation - ALL PARTS, IEC 61850:2023 SER | IEC Webstore | LVDC. Available at: https://webstore.iec.ch/publication/6028 (Accessed: 18 December 2023).

[3] IEC 61850-7-410:2012+AMD1:2015; Consolidated version Communication networks and systems for power utility automation - Part 7-410: Basic communication structure - Hydroelectric power plants - Communication for monitoring and control, IEC 61850-7-410:2012+AMD1:2015 CSV | IEC Webstore | electricity, water management, smart city, LVDC. Available at: https://webstore.iec.ch/publication/23693 (Accessed: 18 December 2023).

[4] IEC 62325-301:2018; Framework for energy market communications - Part 301: Common information model (CIM) extensions for markets, IEC 62325-301:2018 | IEC Webstore. Available at: https://webstore.iec.ch/publication/31487 (Accessed: 18 December 2023).

[5] IEC 61970-301:2020+AMD1:2022 CSV; Energy management system application program interface (EMS-API) - Part 301: Common information model (CIM) base (2022) IEC 61970-301:2020+AMD1:2022 CSV | IEC Webstore. Available at: https://webstore.iec.ch/publication/74467 (Accessed: 18 December 2023).

[6] IEC 61968-100:2022; Application integration at electric utilities - System interfaces for distribution management - Part 100: IEC implementation profiles for application integration, IEC 61968-100:2022 | IEC Webstore. Available at: https://webstore.iec.ch/publication/67766 (Accessed: 18 December 2023).

[7] IEC 62443-2-1 (2010), INTERNATIONAL ELECTROTECHNICAL COMMISSION, NDUSTRIAL COMMUNICATION NETWORKS – NETWORK AND SYSTEM SECURITY – Part 2-1: Establishing an industrial automation and control system security program, Available at: https://webstore.iec.ch/publication/7030/ (Accessed: 18 January 2024).

[8] IEEE Standards Association (no date) IEEE SA - IEC/IEEE International Technical Guidelines for smart hydroelectric power plant, IEEE Standards Association. Available at: https://standards.ieee.org/ieee/63198-2775/7277/ (Accessed: 18 December 2023).

[9] IEC 61970-302:202; Energy management system application program interface (EMS-API) - Part 302: Common information model (CIM) dynamics | IEC Webstore. Available at: https://webstore.iec.ch/publication/68152 (Accessed: 18 December 2023).

[10] BHAT, R. and HUANG, J. (no date) 'Thinking Inside of the Box: Using Autodesk BIM Technology for Hydropower Projects', UT20796. Available at: https://www.autodesk.com/autodesk-university/class/Thinking-Inside-Box-Using-Autodesk-BIM-Technology-Hydropower-Projects-2016#handout (Accessed: 20 December 2023).

[11] https://www.epa.gov/sites/default/files/2015-11/documents/wqx_factsheet.pdf

[12] Horsburgh, J.S. et al. (2016) 'Observations data model 2: A community information model for spatially discrete Earth Observations', Environmental Modelling & Software, 79, pp. 55–74. doi:10.1016/j.envsoft.2016.01.010.