



MARCLAIMED

Requirements and global architecture v1

Deliverable D2.1

WP2

Author: Luc GASSER

Date: 03/05/2024



**Funded by
the European Union**

GRANT AGREEMENT NUMBER	101136799		
ACRONYM / FULL TITLE	MARCLAIMED: integrated Decision Support Tool for Reliable and Affordable Application of Managed Aquifer Recharge with Alternative Water Resources in River Basin and Drought Management Plans		
START DATE	01/02/2024	DURATION	36 months
END DATE	31/01/2027		
PROJECT URL	https://marclaimed.eu/		
DELIVERABLE	D2.1		
WORK PACKAGE	WP2		
CONTRACTUAL DATE OF DELIVERY	31/07/2024		
ACTUAL DATE OF DELIVERY	26/07/2024		
NATURE	R	DISSEMINATION LEVEL	PU
LEAD BENEFICIARY	EGM		
RESPONSIBLE AUTHOR	Luc GASSER		
CONTRIBUTIONS FROM	Luc Gasser (EGM) Franck Le Gall (EGM) Noureddine Bouzidi (USAL) Sara Espinosa Martinez (CETAND) Lupicinio Garcia Ortiz (CETAND) Miquel Tome Carreño (CETBCN) María Guerrero (CETBCN) Alberto Couce Rodriguez Luz Loaiza Tabares		
ABSTRACT	This document is the first version of the Requirements and global architecture for the IDST of the MARCLAIMED project, it is being produced as part of task “T2.1: Co-design of the MARCLAIMED IDST”, which is part of WP2 - MARCLAIMED IDST: Integrated Decision Support Tool for MAR recommendations and early warning alert system.		

The objective of this deliverable is to present the information that has been gathered during the co-design of the IDST: the demo site requirements, for each demo site as well as the tools requirements for each of the 6 MARCLAIMED tools. Based on those requirements, a first version of a global architecture for the MARCLAIMED IDST is presented, with a description of the functionalities of each module.

Disclaimer

Any dissemination of results reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

© MARCLAIMED Consortium, 2024

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. Reproduction is authorised provided the source is acknowledged.

Copyright message

TABLE OF CONTENTS

EXECUTIVE SUMMARY	7
1 Introduction	8
2 Demo sites requirements	9
2.1 Introduction	9
2.2 DS1 (Barcelona, Spain)	9
2.3 DS2 (Comporta, Portugal)	21
2.4 DS3 (Breezand, The Netherlands)	29
3 Tools requirements	33
3.1 MARVIS	33
3.2 PREDMAR	34
3.3 MARAISER	36
3.4 OASIS FLOW	38
3.5 MARINSURE	40
3.6 RECOVER	42
4 IDST global architecture	43
4.1 IDST Data core	44
4.2 Data models	44
4.3 Data collection connectors	46
4.4 Digital Twin Orchestrator and Tools management adapter	47
4.5 IDST user interface	52
4.6 Deployment	52
5 CONCLUSIONS	56
REFERENCES	57

ACRONYMS:

ABM (Agent-based modelling),
AI (Artificial Intelligence),
API (Application Programming Interface),
ASR (Aquifer Storage and Recovery),
ASTR (Aquifer Storage, Transfer and Recovery),
AWARE (Available WATER REMaining per area),
AWR (Alternative Water Resources),
AWRSS (Alternative Water Resources Supply Systems)
BM (Business Model),
CDP (Communication and Dissemination Plan),
CoP (Community of Practice)
DMP (Data Management Plan)
DSS (Decision Support System)
DT (Digital Twin)
DWTP (Drinking Water Treatment Plant)
EU (European Union)
EC (European Commission)
HLEG AI (High Level Expert Group on Artificial Intelligence)
IDST (Integrated Decision Support Tool)
IIP (Intellectual Property)
IPR (Intellectual Property Rights)
KPI (Key Performance Indicator)
MAR (Managed Aquifer Recharge)
ML (Machine Learning)
MLMs (Machine Learning Methods)
PIML (Physics-Informed Machine Learning)
PoMs (Programmes of Measures)
RCP (RepresentativeConcentration Pathways)
RES (RESults)
ROI (Return Of Investment)
SAT (Soil-Aquifer Treatment)
SO (Specific Objectives)
SoA (State of the Art)
SUS (System Usability Scale)
TG (Target Group)
TRL (Technology Readiness Level)
WFD (Water Framework Directive),
WP (Work Package)
WWTP (Waste Water Treatment Plant)

EXECUTIVE SUMMARY

This document is the first version of the Requirements and global architecture for the IDST of the MARCLAIMED project. It is produced early in the project (M6, 31/07/2024) and a second version will be delivered on M24 (31/01/2026). This deliverable is being produced as part of task "T2.1: Co-design of the MARCLAIMED IDST", which is part of WP2 - MARCLAIMED IDST: Integrated Decision Support Tool for MAR recommendations and early warning alert system.

The goal of the deliverable is to present the information that has been gathered during the co-design of the IDST, it covers the demo site requirements, for each demo site as well as the tools requirements for each of the 6 MARCLAIMED tools. From those requirements, a first version of a global architecture for the MARCLAIMED IDST is presented, with a description of the functionalities of each module.

The IDST is centered around a NGSI-LD data core, which will link all the modules together, it will receive data from the data connectors, will send the data to the digital twin orchestrator as well as to the tool management API, that will execute the MARCLAIMED tools. The tools results are all gathered in the data core and can be displayed and compared in the graphical interface.

1 Introduction

Water scarcity is a pressing issue that many regions and countries around the world are suffering from nowadays. This challenge has been exacerbated by the impacts of climate change and the increased demand for water resources regarding socio-economic changing (Schnoor, 2010; Vörösmarty et al., 2000). The MARCLAIMED project adopted two innovative strategies: Managed Aquifer Recharge (MAR) and Alternative Water Resources (AWR) in order to address these challenges and support integrated water management systems. MAR involves the intentional process of increasing groundwater storage, improving water quality, and providing a buffer during periods of drought (Dillon, 2005; Pyne, 2005). Various techniques can be used for MAR, including infiltration ponds and injection wells (Bouwer, 2002). By replenishing groundwater reserves, MAR ensures a reliable water supply during dry seasons and periods of water scarcity (Gale, 2005). On the other hand, AWR encompasses the use of non-traditional water sources. These include reclaimed water, secondary treated water, and agricultural drainage water that are reused for MAR within the framework of the MARCLAIMED project (Angelakis et al., 1999; Lazarova & Bahri, 2005). The aim of utilizing AWR in conjunction with MAR is to provide additional water resources and enhance the resilience of water supply systems against the challenges posed by water scarcity and climate change (Bixio et al., 2005).

The ongoing research in the field of DSS for aquifer recharge solutions is focusing on developing more accurate and efficient independent tools that can aid in the decision-making process. These are related to integration of multiple data sources, improving predictive algorithms as physical models or increasingly using ML and AI, developing multi-criteria optimization algorithms or providing tools for effective collaboration between different stakeholders (water managers, policy-makers, and local communities) to ensure that recharge solutions are socially acceptable and sustainable in the long term. Nevertheless, these are different and distinct tools which hardly share data and do not provide a unified approach to users thus minimizing their potential usage and impact.

To tackle this issue, MARCLAIMED aims to provide through its Integrated Decision Support Tool (ISDT), a flexible framework building on meta data models and application programming interfaces (APIs) allowing the interconnection and execution of 3rd party tools under a uniform interface. The digital twin approach proposed by ETSI in its NGSI-LD specification for context information management will be used as a basis and will be expanded using approaches proposed in SWIM257 which works with regional stakeholders developing a shared understanding of water availability and use under different climate change scenarios, providing a MAR resilience indicator. System definition and development is handled within an agile co-creation process to contribute to users' acceptance and solution effectiveness. The results of this co-design are presented in this deliverable.

2 Demo sites requirements

2.1 Introduction

The implementation of MAR and AWR comes with various challenges. One major concern is ensuring the quality of reclaimed and treated water, which must meet stringent regulatory standards to prevent contamination of groundwater supplies (NRMMC-EPHC-NHMRC, 2009). Public acceptance also poses a significant challenge. Stakeholders have reservations about using alternative water sources, such as reclaimed and treated water, for aquifer recharge (Po et al., 2003). Building trust and sharing knowledge about the safety and benefits of these strategies are essential for their implementation (Dolnicar & Schäfer, 2009). Furthermore, the infrastructure required for MAR and AWR systems demands substantial technical expertise. Constructing and maintaining infiltration ponds, injection wells, and treatment facilities require significant financial investments (Herman et al., 2012). Moreover, proving the economic feasibility of MAR with AWR under current and plausible future socio-economic scenarios is another challenge that should be tackled by conducting economic analyses and demonstrating the long-term economic feasibility of these strategies (Foster et al., 2002).

The MARCLAIMED project aims to integrate MAR and AWR strategies in River Basin and Drought plans. This depends on comprehensive data collection and analysis which includes gathering detailed hydrogeological data to understand the characteristics of aquifers, monitoring water quality to ensure safety, analyzing water usage statistics, and considering climate information to anticipate future water availability (Scanlon et al., 2006). Key stakeholders in the MARCLAIMED project from the different demo sites include policymakers, beneficiaries, public, academic institutions, civil society organizations and end-users such as farmers and industrial entities who rely on water resources (Pahl-Wostl et al., 2007). These stakeholders play essential roles in co-designing, co-develop, and co-evaluate AWR strategies for the MARCLAIMED project.

2.2 DS1 (Barcelona, Spain)

2.2.1 Context and general description

The Spanish demo site is located in the Mediterranean coastal aquifer region of Barcelona, northwest Spain. It employs a combined MAR scheme using Aquifer Storage Transfer and Recovery and Aquifer Storage and Recovery techniques. The primary goal is to enhance the availability and quality of groundwater resources to address the region's water scarcity issues. The site utilizes reclaimed water infiltrated through ponds and a mix of surface water combined with reclaimed water injected via deep wells.

2.2.2 Why this demo site

Barcelona's Mediterranean climate is characterized by severe recurrent droughts and significant water scarcity. The Baix Llobregat aquifer, the primary water source for the region, is highly vulnerable to pollution from industrial and agricultural activities. This demo site was chosen to mitigate these challenges by optimizing groundwater resources through MAR and reducing dependency on surface water sources like the Ter River.

2.2.3 Objectives and impact

The users of the IDST will be the stakeholders from the Spanish Demo site, as presented in the following table (Initial results from T3.1 and T3.2).

Specific demo site stakeholder	Role and competence related to MARCLAIMED	Needs in relation to IDST-MARCLAIMED	Stakeholder group	Scale
Catalan Water Agency (ACA)	Public company in charge of water planning and management. Its action plan guarantees, now and in the future, the supply, availability of water and its quality at source. It also promotes the sanitation of waste water and the protection and conservation of water bodies and associated ecosystems.	Improved management of scarce water resources towards climate resilience and sustainability; maintained long-term, safe and efficient water supply.	Policy makers	Regional

Metropolitan area of Barcelona (AMB)	Public administration with different types of competencies, among which water cycle and water resources, supply, wastewater treatment, reuse and flood prevention.	Ensured and improved quality, continuity and efficiency of water services, considering economic and financial sustainability and the principles of social and environmental responsibility.	Policy makers	Metropolitan
Public Health Agency of Catalonia (ASPCAT)	Authority responsible for the promotion and protection of health, the prevention and management of epidemiological and food alerts, and occupational health.	Guarantee the water supply necessary to safeguard the health of the population, monitoring its quality.	Policy makers	Regional
City councils (<i>Barcelona, Sant Vicenç dels Horts, Molins de Rei...</i>)	Ensures water security and integrity of the environment on a local level.	Guarantee the water supply necessary to safeguard the health of the population.	Policy makers End-users	Local

Consorti del Delta del Llobregat	Protects and manages Llobregat Delta's natural spaces, monitors the development of activities to ensure compatibility with the ecological functions and services provided by these natural spaces, while facilitating coordination across relevant public authorities.	Improved protection and management of river natural resources for environmental integrity and continuity of water services; improved data and knowledge on environmental impacts.	Public Entity	Local
Associació Catalana de Comunitats de Regants (ACATCOR)	Grouping of all communities and collectivities of irrigators, central water user unions, unions, irrigation water concessionaires, etc. of Catalonia with the purpose of advancing savings in its operation and exchange knowledge and practices for the improvement of irrigated areas in the region.	Improved water management and related costs in Catalunya; improved water availability and quality; water supply security; modernization of activities to increase efficiency; equitable distribution of water resources.	Public Entity	Regional

Grup de Recerca en Ciència i Tecnologia de la Sostenibilitat (UPC)	A research organization focused on advancing sustainability through innovation, including developing technical and conceptual tools such as modeling and simulation of complex systems.	Improved knowledge in innovative and sustainable solutions; Integrate the project outcomes in future research projects	Academia; R&D	
Polytechnic University of Catalonia (UPC)	Contributes to building a sustainable and fair world through research, technology transfer and knowledge dissemination, by producing engineering, architecture, science and technology graduates.	Analyze the project and follow its progress to acquire useful knowledge and be able to export sustainable practices in other contexts.	Academia	Local
University of Barcelona (UB)	One of its missions is social commitment and, for that, it has assumed global challenges, such as poverty, inequality, climate change, environmental degradation, or the fight for prosperity, peace and justice.	Analyze the project and follow its progress to acquire useful knowledge and be able to export sustainable practices in other contexts.	Academia; R&D	Local

Research Institutes (<i>ICRA, IGME-CSI, IDAEA-CSIC, ICGC</i>)	Public research institutions that conduct R&D activities, offer consultancy (including to the government), and provide technical and academic expertise in several scientific areas	Integrate the project outcomes in future research projects	Academia; R&D	Regional
Grup d'Hidrologia Subterrània (GHS)	Research, teaching and knowledge transfer in the field of hydrogeology and geochemistry, specifically applied to groundwater resource assessment, aquifer management, wetland studies, marine intrusion and underground waste and CO2 extraction.	Improved knowledge in groundwater resource management; integrate results and further research based on the project outcomes	Academia; R&D	

Water Users Community from Llobregat Valley and Delta (CUADLL)	Entity ascribed to the Catalan Water Agency (ACA) with functions of distribution, administration, protection, and preservation of surface and groundwater. Among delegated powers include recharging basins and improving and expanding this infrastructure. The CUADLL represents all water users.	Improved water infrastructure and management of Llobregat groundwater; improved water availability and quality; water supply security; modernization of activities to increase efficiency; equitable distribution of water resources.	Public Entity	
Irrigation community (Comunitat de regants de la Palma, Comunitat de regants de Corbera de Llobregat, Comunitat de regants bassa de Ferrer i Mora...)	Public law corporations attached to the Catalan Water Agency. Their purpose is the collective administration of the water use rights that have been granted or recognized.	Increase water availability and efficiency for irrigation activity; improved use of alternative (MAR) methods and its monitoring to increase efficiency; greater resilience to water stress and climate change.	End-users	Local

Water utilities (Aqualia, AICSA, Aigües del Prat, Sorea, Sant Vicenç dels Horts)	Carries out the operation and management of water services in the respective municipality, which includes catchment, treatment, distribution, and operation of wastewater treatment stations.	Ensured and improved quality, continuity, and efficiency of water services, considering economic and financial sustainability; improved use of alternative methods such as MAR to increase infrastructure efficiency and achieve environmental sustainability goals such as the increased use of treated wastewater; improved management of demand peaks and decreased stress in natural systems; fair distribution of water.	Water Utility	Local
---	--	--	------------------	-------

Catalan Water Partnership	Non-profit strategic association, formed by companies and knowledge centers that operate in the sustainable water use sector with the objective of improving the global competitiveness of its associates.	Developing innovative and sustainable solutions for efficient water use; secure water supply; improved data and knowledge on environmental impacts and monitoring of MAR systems; improved water availability; expansion of water infrastructure and management efficiency for its associate's activities.	Partnership of companies and research centers	Regional
Companies extracting water from the aquifer (<i>Damm, Cementos Molins, Nestle and Celsa</i>)	Companies carrying out industrial activities intensive in water consumption. Both companies own wells in the Vall Baixa and Delta Llobregat aquifer. Both companies promote the implementation of AWR infrastructures and solutions and solutions.	Increase water availability and efficiency for industrial activity; water supply security; expansion of MAR infrastructure; improved use of alternative (MAR) methods and its monitoring to increase efficiency; reduced water footprint; greater resilience to water stress and climate change.	End-users/ Industry	Local

Agrupacions de Defensa Vegetal (ADV) (del Baix Llobregat d'Horta and del Delta del Llobregat)	Non-profit private entities composed by farmers with the objective of collaborating with competent public administrations in integrated pest management.	Innovative solutions for sustainable agricultural practices and water use; improved data and knowledge on environmental impacts and be able to export sustainable practices in other contexts	Civil Society	Local
Water-related Civil Society Organizations (<i>Associació catalana amics de l'aigua, Agua segura project</i>)	Organizations dedicated to promoting and implementing initiatives and projects related to water, sanitation and hygiene (WASH) with the aim of generating knowledge and a positive impact at the basin, communities and businesses level in Catalonia. These organizations promote solutions and increase awareness on water-related issues in the territory.	Improved data and knowledge on environmental impacts and innovative solutions; Improved management of scarce water resources towards climate resilience and sustainability; maintained long-term, safe and efficient water supply.	Civil Society	Regional
Workers (<i>CUADLL, Aigües de Barcelona</i>)	Workers potentially impacted by the project and its results.	Capacity building; increased efficiency in work activities	Civil Society	Local

Metereological Service of Catalonia (METEOCAT)	The Climatology area participates in research projects related to climate change, and also to dissemination of climate knowledge across various fields and to diverse groups.	Knowledge sharing	Public Entity	Regional
--	---	-------------------	---------------	----------

2.2.4 Infrastructure description and available datasets

The infrastructure scheme for the Sant Vicent dels Horts (SVH) MAR system (infiltration ponds) is presented below.

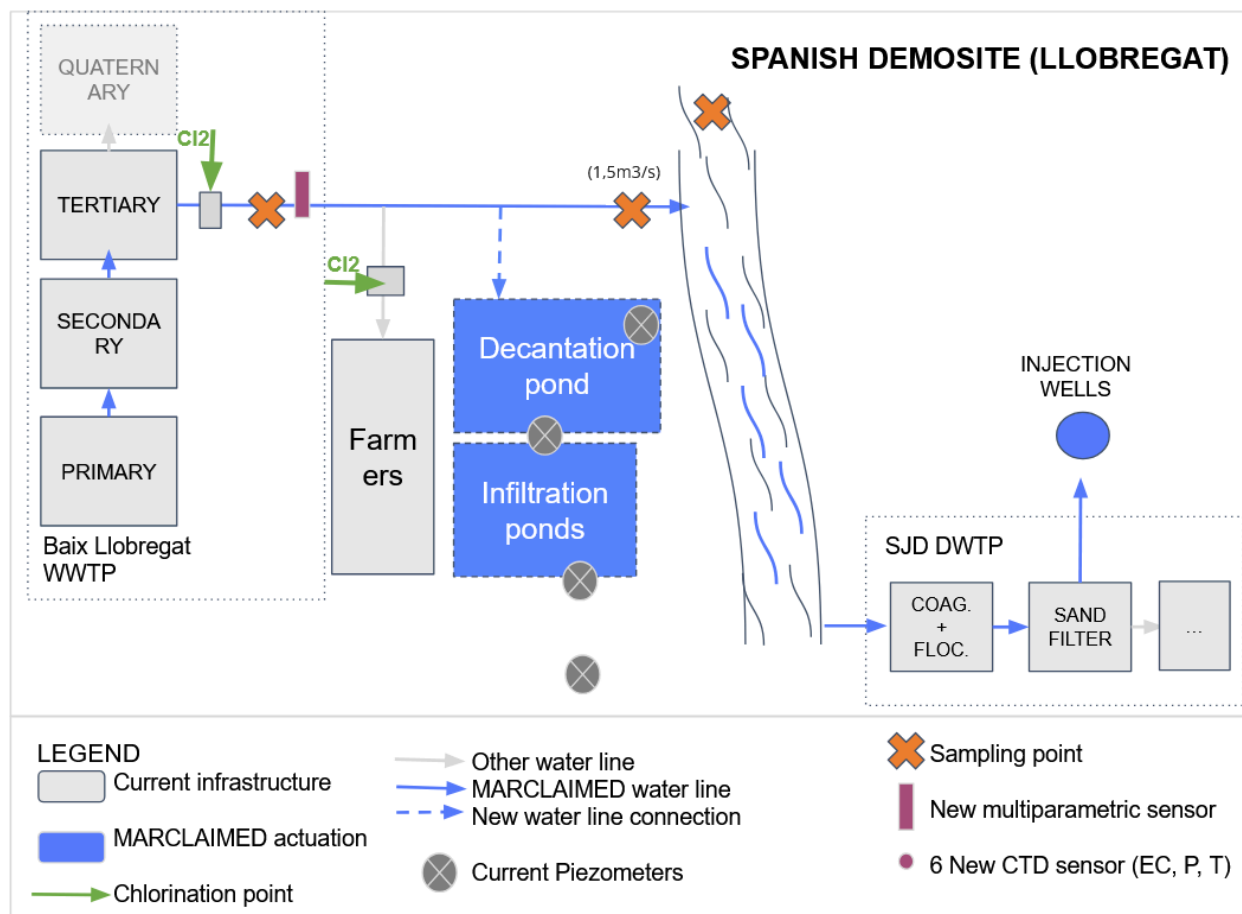


Figure 1: DS1 infrastructure (SVH infiltration ponds) diagram

The main elements are the followings:

- Infiltration pond with an infiltration capacity of 3900 m³/day.
- Injection well with an injection capacity of 1340 m³/day for a mixture of water sources.
- Water Treatment Plants: Sant Joan Despí Drinking Water Treatment Plant and El Prat Water Reclamation Plant.
- Monitoring Systems: Track water quality, recharge rates, and piezometric levels.

The available data are historical data, including hydrogeological data, water quality data, climatic data, and usage data and economic data have been provided by the demo site leaders and shared with the tools' developers to be used as essential inputs for all the tools of MARCLAIMED. Additional data will be generated during the project.

The scheme of the deep injection MAR systems presented below:



Figure 2: DS1 infrastructure (P18 deep injection MAR system)

2.3 DS2 (Comporta, Portugal)

2.3.1 Context and general description

The Portuguese demo site is located in Comporta, part of the Atlantic coastal aquifer system. The region is characterized by high risks of water scarcity and extreme droughts. The site utilizes a MAR scheme with Soil-Aquifer Treatment (SAT) using secondary treated wastewater, enhancing groundwater recharge and improving discharged water quality.

2.3.2 Why this demo site

This demo site faces significant water management challenges due to frequent droughts. The site's proximity to the sensitive Sado estuarine ecosystem, classified as a Nature Reserve, further requires sustainable water management practices, not allowing direct discharge of treated effluents into surface water bodies. This site demonstrates the effectiveness of MAR and AWR in addressing these issues and improving regional water security.

2.3.3 Objectives and impact

The users of the IDST will be the stakeholders from the Portuguese Demo site, as presented in the following table (Initial results from T3.1 and T3.2).

Specific demo site stakeholder	Role and competence related to MARCLAIMED	Needs in relation to IDST-MARCLAIMED	Stakeholder group	Scale
Portuguese Environmental Agency - APA (including the Alentejo regional offices)	River Basin Administration with monitoring and water use licensing responsibilities, which may include MAR licensing once legal framework is established	Improved water resources management; demonstrate MAR capabilities and safety	Water authority, policy maker	National, regional

Águas Públicas do Alentejo (AgdA-AdP)	Operation and management of bulk water supply under a public partnership regime. Provides water for public supply to the region's municipalities and local entities and ensures wastewater treatment	Ensure that MAR has a set of safe methods that allows for an effective valorization of treated wastewater effluents and improvement of water availability	Water supply and wastewater treatment	Regional
Coordination and Regional Development Commission for the Alentejo region (CCDR-Alentejo)	Coordinates various sectorial policies at regional level regarding territorial planning/land use. Provides financing mechanisms to support local authorities and associations with the objective of achieving regional economic and social development	Promote MAR implementation through public financing mechanisms	Policymaker	Regional

Portuguese Farmers Confederation (CAP)	Promotes the development of technical analysis, specialized studies or strategies for improved irrigation practices, promoting water and energy efficient use; represents the interests of agricultural sector and rural communities	Promote alternative and safe water management practices such as MAR that may provide clear benefits to environmental services	End-users; Sectorial representative	National
Municipalities and Alentejo Intermunicipal Communities (AMAL, CIMAA, CIMAC, CIMBAL)	Ensures water distribution at local level; Promotes environmental sustainability practices and increases water user's awareness to water related issues; implementation of local water management strategies	Benefits from alternative water management policies and practices at local level, particularly from environmental services protection for local social and economic development	End-users	Regional, local

Local associations for farming irrigators and beneficiaries (e.g., ABORO, ABROXO, ARBCAS)	Manages water distribution infrastructure for irrigation purposes at local level; manages water allocation and settles disputes related to water availability	Improved water availability and quality due to integrated water management policies at regional and local level	Irrigators associations	Regional, local
Alqueva Infrastructure and Development Company (EDIA)	Public company that manages the Alqueva Multipurpose Project which includes a dam and water distribution infrastructure; one of the largest water provider for irrigation in the Alentejo	Improved and complementary water management practices that contribute for increased availability; ensure that MAR does not affect water quality and removes the pressure over the surface water resources	Irrigation water supply	Regional, local

Portuguese Water Partnership	Encourages an effective link between professionals, institutions and companies in order to project the knowledge and skills of the Portuguese water sector in the world, and to catalyze opportunities in international markets and in the area of cooperation within the framework of the development of sustainable projects in line with the Sustainable Development Goals	Promotes innovation in the water sector, facilitating cooperation between companies and research centers; takes an important role in making knowledge in the subject of water resources management available to all possible interested water companies and institutions, both at national and international scales (e.g. Portuguese-speaking countries)	Partnership of water-related companies and research centers	National
Évora University (UE)	Public higher education and research institution (and foundation), with technical and academic expertise	Integrate the project outcomes in future research projects	Academia; R&D	Regional

	in several scientific areas			
Beja Polytechnic Institute (IPBeja)	Public higher education institution with technical and academic expertise	Integrate the project outcomes in future research projects	Academia; R&D	Regional
National Laboratories (LNEC - civil engineering; LNEG - geological services)	Public research institutions that conduct R&D activities, offer certification and consultancy (including to the government)	Further promote research related activities based on the project outcomes; ensured that innovative practices are considered in policy making	R&D	National
Regional Board of Agriculture and Fisheries (DRAP-Alentejo)	Participates in the formulation and execution sustainable policies in the areas of agriculture, forests, rural development and fisheries, as well as collaborates in food safety and farming activities policies; provides guidelines in improved water usage techniques for farmers	Safer and affordable strategies and policies to secure availability of water for irrigation; improved and safe usage methods of non-conventional water sources; support the integration of innovative practices in policy-making	Agricultural authority; policymaker	Regional

Tourism-related SMEs	Tourism sector companies aiming to implement sustainable and environmentally responsible leisure activities; promote the protection of environmental services and benefit from local "green" policies	Benefit from the implantation of water resources management policies that can improve water quality and availability	End-users, industry	Local
Large end-users from Agricultural Sector (e.g., SOVENA, WESTFALIA)	Farming companies and water users for irrigation; integrate innovative farming practices	Benefit from the implantation of water resources management policies that can improve water quality and availability	End-users, industry	Regional, local
Construction & environmental consultancy companies	Feasibility studies; design and construction; environmental impact assessment; regulatory compliance implementation (water rights, environmental protection laws and construction codes); R&D	Benefit from ready-to-market MAR implementation guidelines for successful replication and business expansion/promotion of such practices	Industry	National, regional, local

2.3.4 Infrastructure description and available datasets

The infrastructure of the demo site is presented below.

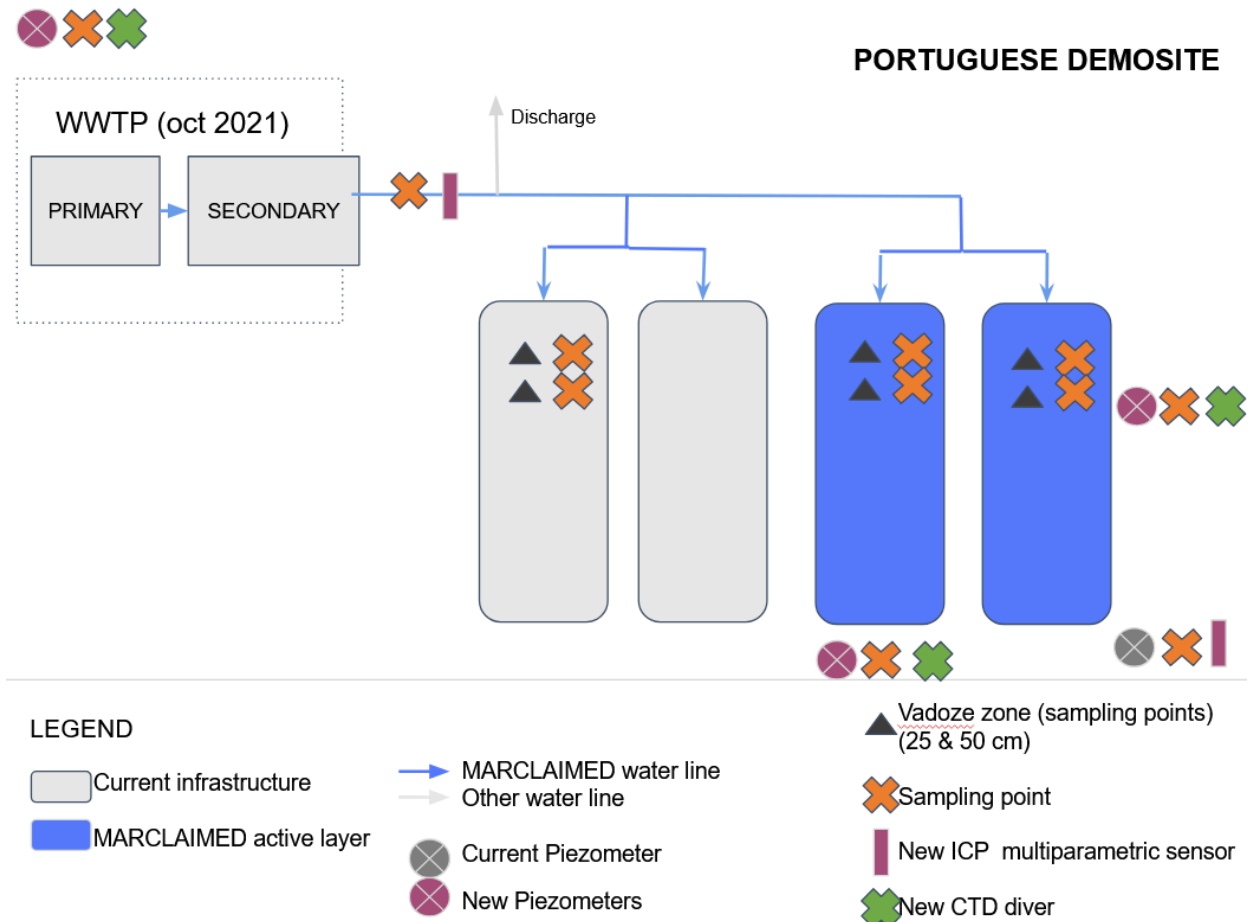


Figure 3: DS2 infrastructure diagram

The main elements are the followings:

- Infiltration Ponds: Four ponds utilizing secondary treated wastewater with a capacity of 190 m³/day.
- Treatment Plants: Comporta Wastewater Treatment Plant with advanced systems.
- Monitoring Systems: Track water quality and infiltration performance.

Historical data, including hydrogeological data, water quality data, climatic data, and usage data and economic data have been provided by the demo site leaders and shared with the tools' developers to be used as essential inputs for all the tools of MARCLAIMED (Work Package 1). Additional data will be generated during the project adding to the existing monitoring facilities three recently (july) installed piezometers and six capsules in the vadoze zone. Three CTD divers and two scan probes were installed for continuous and online data gathering.

2.4 DS3 (Breezand, The Netherlands)

2.4.1 Context and general description

The Dutch demo site is located in Waddenregio, northwest Netherlands, near the Wadden Sea. The site employs an Aquifer Storage and Recovery (ASTR) system using agricultural tile drainage water to address water scarcity and pollution.

2.4.2 Why this demo site

Waddenregio was selected due to its intensive agricultural activities, significant water scarcity challenges, and vulnerability to pollution. The region's economic reliance on agriculture, particularly flower bulb cultivation, underscores the need for sustainable water management solutions.

2.4.3 Objectives and impact

The users of the IDST will be the stakeholders from the Dutch Demo site, as presented in the following table (Initial results from T3.1 and T3.2).

<i>Specific demo site stakeholder</i>	<i>Role and competence related to MARCLAIMED</i>	<i>Needs in relation to IDST-MARCLAIMED</i>	<i>Stakeholder group</i>	<i>Scale</i>
Water Framework Directive (WFD) / EU	WFD aims to achieve good ecological and chemical status in European water bodies.	Prove that MAR benefits water quality and ecological value	Government	EU
Dutch national government	Suffice with WFD Suffice with nitrogen norm	Prove that MAR has a positive effect on nitrogen emission/leakages and water quality when applied to farming systems.	Government	National

Province	Support economic development/sustainability in their province Secure drinking water Ensure ecological value	MARCLAIMED benefits the feasibility of MAR in the region, therefore it supports sustainable agriculture/economic development, and it supports the goals of the Deltaprogramma Zoetwater which is to support freshwater self-sufficiency; Understand how MAR (positively) affects drinking water supplies. How MAR can also be applied as tool for better ecological management (e.g., reduce nutrient leakages to nature, alternative water supply during droughts)	Government	Province / regional
----------	--	--	------------	---------------------

Waterschap (water authority)	Main water management authority. Their role is to provide/manage the fresh water distribution. It is not their duty to supply fresh water, but they do have an obligation to make an effort and do the best they can. Allround improvement of the water quality.	MAR should be proven to benefit these interests. MAR on farms is a very local way to control water levels. In the current pilots, the water is purified before infiltrating, hence this benefits water quality and nature.	Government	Regional / local
Farmers	Invest in MAR Investment opportunity	MAR should support investment	Private / end user	Local

2.4.4 Infrastructure description and available datasets

The infrastructure of the demo site is presented below.

DUTCH DEMOSITE (TEXEL)

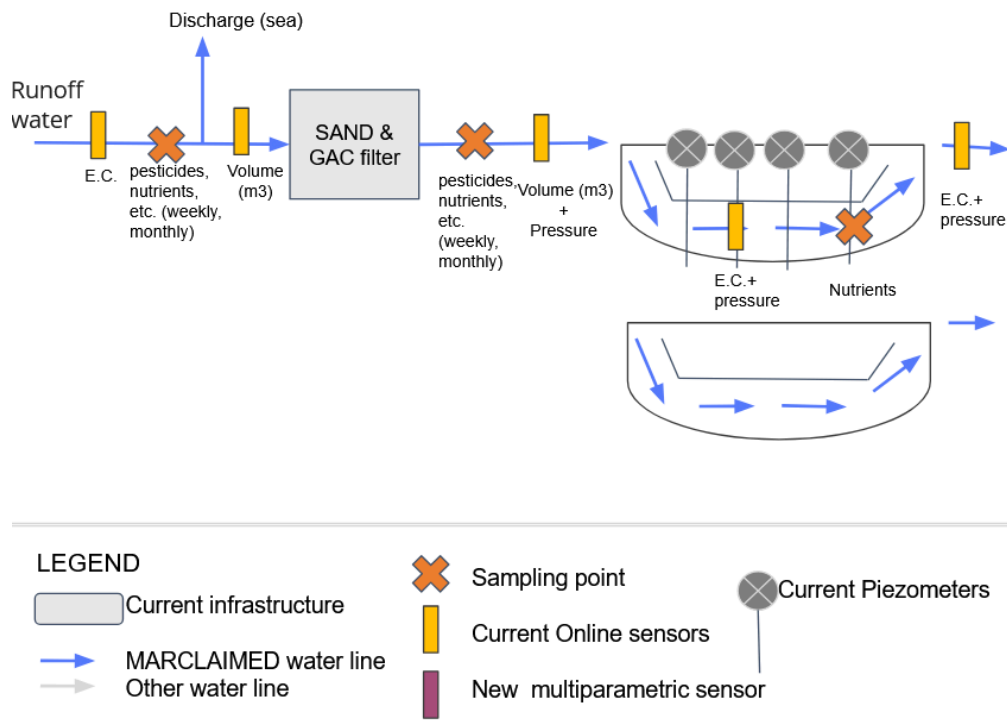


Figure 4: DS3 infrastructure diagram

The main elements are the followings:

- ASTR MAR System: Utilizes agricultural tile drainage water with a capacity of 400 m³/day.
- Abstraction Wells: Ensure a reliable supply of water for agriculture.
- Pre-treatment Facilities: Filtration and biodegradation methods.

Available data are historical data, including hydrogeological data, water quality data, climatic data, and usage data and economic data have been provided by the demo site leaders and shared with the tools' developers to be used as essential inputs for all the tools of MARCLAIMED (Work Package 1). Additional data will be generated during the project.

3 Tools requirements

To give some context on what is to be orchestrated by the API, the 6 MARCLAIMED tools are presented hereafter.

3.1 MARVIS

3.1.1 Purpose of the tool

Virtual sensing tool for near real-time water quality monitoring that will be developed using the data from 2 demo sites (ES and NL) and the fitness of its implementation in the remaining demo site (PT) will be studied after it has been tested in the other demo sites. For this tool, a water quality parameter related to pollution from target emerging organic contaminants will be identified for each demo site. By using the available training data from historical datasets provided by NL and ES demo sites, a Machine Learning model will be trained and validated, resulting in two different virtual sensors (one for each parameter identified in the previous step). Eventually, MARVIS will be integrated in the MARCLAIMED IDST and demonstrated in NL, ES and in PT (as mentioned above).

MARVIS aims to provide near real-time detection of target emerging organic contaminants in AWR by means of virtual sensors, a novel method for environmental monitoring which looks for correlations between the monitored parameter and other parameters of concern. MARVIS virtual sensors will be developed on historical datasets provided by ES and NL demo sites and tested in all 3 demo sites. In the specific case of the PT demo site, an approach based on transfer learning will be adopted during the testing phase since no specific historical data for the site is available for training.

With the use of Machine Learning models, MARVIS will target the inferred near real-time detection of emerging organic contaminants. The data requirements of the Deep Learning approach will be considered by putting additional efforts into data management and data enrichment tasks. From a pure data-driven perspective, the development of the tool will cover the entire lifecycle of ML modeling from target identification to virtual sensor model validation and testing. Results will lead to defining early warning risk indicators (parameters and target values) and automated systems to prevent source contamination by deviating inadequate water for recharge.

Finally, accuracy of the virtual sensing solutions will be tested by deploying them into data pipelines providing unseen data streams (different from those used in the training process). This can be done by comparing the inferred contaminants levels from the Virtual Sensors with actual measurements taken in the field. MARVIS 'proactive approach can prevent damage to the aquifer and minimize potential health risks associated with the use of reclaimed water. The results of MARVIS will feed the MARAISER tool.

3.1.2 Configuration parameters

The IDST will have to access both sensoric data and sampling data (see input data below), so, depending on the demo site, it will need access to the sensor directly, to an intermediate storage (SFTP, FTP, etc.) or to a database where the required data can be queried.

3.1.3 Input data

The virtual sensors will feed on different data sources:

- Sensoric data that provide near-continuous values (flow, turbidity, nitrates, conductivity, etc.).
- Sampling routine data (nutrients, organic matter, physico-chemical parameters, etc.).

3.1.4 Output data

Estimation of the target parameter of interest for each demo site. The result will be either a float estimating the quantity of the parameter or a range indicating threshold values for the presence of the contaminant.

3.1.5 Triggering mode

Continuous monitoring of the target parameter for each demo site. The frequency of the inference will vary depending on the necessities of the demo site, which will be able to decide if the parameter is estimated hourly, daily, weekly, etc.

3.1.6 Deployment model

The result of the work performed during the development of the MARVIS tool will be a trained model that can be delivered in a wide variety of formats.

3.2 PREDMAR

3.2.1 Purpose of the tool

AI-based water availability and demand forecasting tool will be implemented using the data coming from the all demo site (ES, PT and NL), to develop the corresponding functionalities for water availability (stored water in reservoirs, surface water surpluses) and water demand (including groundwater abstraction from NL) forecasting models in short (months) and long term (years). A framework for integrating data from Machine Learning and deterministic models (e.g., meteorological and/or hydrological services) into one coupled tool will be designed, with the final goal of boosting the use of AWR throughout the year, and not only during drought periods. The tool will be integrated in the MARCLAIMED IDST.

PREDMAR aims to establish a sustainable use of the available water resources in a climate change context. PREDMAR will be validated in all 3 demo sites but under different approaches.

In the ES demo site, PREDMAR will predict the quantity of stored water in the main reservoirs of the Baix Llobregat watershed at two different time scales:

- 1) short (3 - 6 months) and
- 2) long-term (decadal predictions based on different climate change projections regarding Representative Concentration Pathways (RCP) 4.5 and 8.5 emission scenarios). Climate change projections will be downscaled and bias-corrected to represent the local climate conditions following a Machine Learning approach.

This approach will compare, for a specific period of time in the past (called 'hindcast'), the downscaled/simulated values and the observed ones in different rainfall weather stations. The predictive approach will evolve towards a hybrid tool by integrating information (stream flows, stored water) from a deterministic hydrological model already developed in an ongoing European project (HEU ICARIA - CETBCN: leader; AIGUES: partner).

PREDMAR will also include a functionality to analyze and predict demographic changes and its relationships with water demands, the predicted stored water will be linked to the 'Drought Management Plan' to anticipate when the system must start incorporating reclaimed water to increase the overall water resilience.

For the PT demo site, PREDMAR will focus on the identification of surface water surpluses at basin/subbasin scales to assess the feasibility of designing new MAR schemes. PREDMAR will be combined with obtained results in an ongoing PRIMA project (AGREEMAR, LNEC partner) in which a daily sequential soil-water budget hydrological model has been calibrated.

Finally, for the NL demo site, PREDMAR will be replicated to predict the groundwater to be pumped in the next growing season, using time series of electrical conductivity, water volumes of harvested water, inflow, and outflow of each individual well, and irrigation volumes.

The results of PREDMAR will feed the following tools: MARAISER, OASIS FLOW, MARINSURE and RECOVER.

3.2.2 Configuration parameters

The IDST will have to access data (see input data below), so, depending on the demo site, it will need access to an intermediate storage (SFTP, FTP, etc.) or to a database where the required data can be queried.

3.2.3 Input data (example of DS1, will be different for DS2 and DS3)

- Data on the volume discharged into the reservoirs of:
 - La Baells
 - La Llosa del Cavall
 - Sau
 - Susqueda
- Data on contributions to the reservoirs of:
 - La Baells
 - La Llosa del Cavall
 - Sau

- Susqueda
- Reclaimed water flow data
 - Recharging system (ponds)
 - Upstream river

3.2.4 Output data

Time series results for each site.

3.2.5 Triggering mode

Continuous monitoring of the target parameter for each demo site. The frequency of the inference will vary depending on the necessities of the demo site, which will be able to decide if the parameter is estimated hourly, daily, monthly, etc.

3.2.6 Deployment model

The result of the work performed during the development of the PREDMAR tool will be trained models that can be delivered in a wide variety of formats.

3.3 MARAISER

3.3.1 Purpose of the tool

MARAISER is a risk management tool for Managed Aquifer Recharge focused on the use of Alternative Water Resources as recharged water. It is composed of two modules: Module 1 (short-term time perspective): to improve the “day-by-day” system operation for both MAR schemes infiltration ponds and deep injection wells, and Module 2 (long-term time perspective): to optimize the volume to be recharged according to the water resources availability and demand forecasting.

Module 1 delivers a set of environmental, health, and performance risk indicators based on water quality and hydrodynamic data, improving the operation efficiency in terms of water quality monitoring and MAR while providing operational risk assessment. With this purpose, the MARAISER is fed with the data registered by sensors (online/offline commercial sensors) together with MARVIS strategically located control points of the MAR systems, as well as with data based on the laboratory analysis from the water sampling campaigns of recharge water, soil, and groundwater. Nevertheless, the selection of variables to be monitored may vary among demo sites as it will be based on regional policies and regulation, and will follow previously defined instructions to ensure the health of the aquifer. MARAISER provides a time series evolution with the most relevant statistical measurements. Based on the thresholds/setpoints established, a set of risk indicators, such as microbial, chemical, clogging, salinization, water exploitation and the recovery efficiency (ratio between the extracted volume and the infiltrated volume), will be calculated and lead to alarms with recommendations for actions to mitigate the potential risk in the MARCLAIMED IDST.

Module 2 integrates the forecasted data series of water resources availability and demand, and the recovery efficiency indicator. This will allow to optimize the volume recharged by

providing design adjustment recommendations through the MARCLAIMED IDST, to achieve the optimal performance allowing the planification at long term of improvements in the infrastructure design and in the operation management strategies.

3.3.2 Configuration parameters

Kind of data:

- Water quality data (csv): EC, pH, T, major ions, metals, CECs, microbiology.
- Hydraulic parameters (csv): Flow rate and pressure.
- Hydrodynamic groundwater parameters (csv): piezometric levels, pumping flow rate.
- GIS layers for visualization: vector and raster.

Configuration:

- Sensors to be connected.
- Database to connect to, to read and save observed data and analyzed results

3.3.3 Input data

Module 1:

- Commercial sensors (online and offline): water quality monitoring data (surface, alternative water resources and groundwater), hydraulic parameters (MAR system and groundwater), both of them from the critical control points (CCP) defined.
- Virtual sensing monitoring of emerging contaminants from MARVIS tool.
- Laboratory analysis: water quality data from the CCPs.
- Physico-chemical In situ measurements: collected during fieldwork sampling campaigns from the CCPs.

Module 2:

- Water resources availability and demand forecasting from PREDMAR tool.
- Recovery efficiency indicator.

3.3.4 Output data

- Time series evolution with the most relevant statistical measurements.
- Health (microbial), environmental (salinization and water exploitation) and operational (clogging and recovery efficiency) indicators.

3.3.5 Triggering mode

Execution flow chart of different setpoints depending on the risk, the indicator and the parameter (based -if any- in legal thresholds) to establish different status of alert. The different status of alert provides recommendations to mitigate the identified risks.

This tool runs on with the maximum availability frequency of data.

3.3.6 Deployment model

MARAISER Tool is not an actual model, but the digitalization of the flowchart. This digitalization will be done in the form of scripts deployed within the IDST.

3.4 OASIS FLOW

3.4.1 Purpose of the tool

The OASIS FLOW tool is a modeling tool that aims to integrate the water footprint calculations for the different economic sectors within a specific territory, with the water footprint related with the aquifer recharge associated with the same territorial unit. All of these use downscaled AWARE (Available Water Remaining) indicators to ensure the accuracy and specification of the process.

3.4.2 Configuration parameters

Consumptive water uses determination.

Crop characterization.

Livestock characterization.

Tourism availability and characteristics.

Hydrological characterization of the aquifer/s involved.

Determination and characterization of the discharges and effluents.

3.4.3 Input data

MUNICIPAL CONSUMPTION

- Abstractions
- Own Abstractions for Municipal Supply
- Own Abstractions - Industrial
- Other Water Saving Measures (Municipality)
- Saved Volume
- Population Grey Water (in case of discharge to the sea, it counts as indirect grey water)
- Treated Flow
- C max
- C nat
- C effl
- V effl
- Effluent Grey Water
- Grey Water WWTP
- % WWTP in Municipal Water Treatment
- Grey Water Associated with the Municipality

AGRICULTURE *

- Use and Overload and Value of Used Surface
- Conifers
- Conifers Associated with Other Broadleaf Trees
- Herbaceous Crops under Irrigation
- Fruit Trees under Irrigation
- Dryland Fruit Trees
- Vegetable Gardens or Forced Crops
- Unproductive
- Dryland Cultivation
- Shrubland
- Shrubland Associated with Conifers
- Shrubland Associated with Conifers and Broadleaf Trees
- Pastureland-Shrubland

Agricultural uses should be adapted depending on specific crops for the designated case study

LIVESTOCK *

- Type of Municipal Farm
- Percentage of Exploitation for Each Animal Type and Farm
- Water Footprint of Each Animal Type at the Individual Level
- Water Consumption of Each Animal
- Fraction of Concentrated Feed in Total Dry Food for Each Animal Type, Production System, and Region

Livestock uses should be adapted depending on specific animal production products for the designated case study

TOURISM

- Hotels Average Consumption per Person per Day
- Water Use per Tourist per Day
- Occupancy Survey for Different Types of Tourist Accommodations Available

3.4.4 Output data

1. Water Footprint for each of the different economic sectors taken into account during the design of this tool.
2. Water Footprint associated with the MAR (Managed Aquifer Recharge).
3. Combined assessment of the water footprint calculations taken into account the effects of the MAR

3.4.5 Triggering mode

Scenario-Based Analysis: Runs based on the selection of scenarios and input data characterization for each case study , based on the determination of the configuration parameters mentioned above.

3.4.6 Deployment model

The deployment method is yet to be discussed between developers and the research team in order to provide the most optimal solution to access and review the results provided by this tool.

3.5 MARINSURE

3.5.1 Purpose of the tool

MARINSURE is an ensemble forecasting tool to assess the socioeconomic performance, both at a local (micro-) and economy-wide level (macro-economic), of MAR with AWR strategies under the uncertainty of future plausible scenarios, such as price changes, climate chanpacts or socio-political instability, among others. MARINSURE will assemble, calibrate, and validate an actionable modeling suite of socioeconomic models using a protocol-based modular approach that will allow to measure the economic value of the water security provided by MAR to different present and future uses, mitigating future supply-demand imbalances. Multiple models of microeconomic (mathematical programming models such as Positive Mathematical Programming, Positive Multi-Attribute Utility Programming) and macroeconomic systems (input-output, computable general equilibrium models) will be used. Data will be gathered from relevant data sources including micro- and macroeconomic statistics and will consider the outputs from PREDMAR (changes in water availability), OASIS FLOW (impact of the water footprint), and RECOVER (cost recovery capacity), accounting, parameterizing, calibrating, and validating the models at each system/module level, for each demo site. The upshot will be an actionable tool that can be readily used to formulate and address locally relevant problems related to water resilience by providing information on changes in employment and Gross Value Added.

3.5.2 Configuration parameters

- Land Use scenarios and Crop Portfolio
- Prices and Yield changes
- Costs of crop production, operation of water treatment, and aquifer recharge
- Employment, Subsidies and Insurances

- Water Availability and demand

3.5.3 Input data

- Economic Statistics: Local and economy-wide statistics related to prices, costs, yields, subsidies, and insurances.
- Employment Data: Statistics on employment trends and patterns.
- MARCLAIMED Tools' outputs: Incorporates outputs from PREDMAR (changes in water availability), OASIS FLOW (impact of the water footprint), and RECOVER (cost recovery capacity).

3.5.4 Output data

- Performance Metrics: Provides microeconomic and macroeconomic performance metrics.
- Water Security Value: Assesses the economic value of water security for various uses.
- Employment Changes: Forecasts changes in employment levels.
- GVA Changes: Evaluates changes in Gross Value Added.

3.5.5 Triggering mode

- Scenario-Based Analysis: Runs based on the selection of scenarios and input data, allowing users to explore various plausible scenarios and their economic implications.

3.5.6 Deployment model

- Mathematical Optimization: Built using the General Algebraic Modeling System (GAMS) for high-level mathematical optimization.
- Multi-Language Integration: Utilizes R and Python for sensitivity analysis, models coupling, and result analysis.

3.6 RECOVER

3.6.1 Purpose of the tool

The RECOVER tool aims to support the tariff setting to recover the costs incurred to execute the generation and use of reclaimed water. The establishment of clear unit costs carried by the different stakeholders (operators, regulators, public administrators) will allow the further development of the reclaimed water supply and the effective implementation of tariffs to ensure cost recovery.

3.6.2 Configuration parameters

- Construction costs
- Operation and maintenance costs
- Distribution capacity and economies of scale
- Consumption and potential demand of each stakeholder
- Opportunity cost (restrictions and other water resources)
- Regulation: subsidies, restrictions and requirements

3.6.3 Input data

- unit costs
- demand
- supply
- costs of alternative resources

3.6.4 Output data

- Optimal tariff for cost recovery
- Sensitivity analysis

3.6.5 Triggering mode

- Scenario-Based Analysis: Runs based on the selection of scenarios and input data, allowing users to explore various plausible scenarios and their economic implications.

3.6.6 Deployment model

The deployment method is yet to be discussed between developers and the research team in order to provide the most optimal solution to access and review the results provided by this tool.

4 IDST global architecture

The MARCLAIMED IDST will provide accurate and reliable data and information that can inform decision-makers related to MAR with AWR to support a proactive and adaptive approach to AWR management, taking into account a range of potential scenarios and contingencies. It will facilitate collaboration and communication among stakeholders with diverse interests and perspectives, in order to promote more effective and equitable decision-making and increase public acceptance and decision-makers' reliability. It will promote transparency and accountability in decision-making, by providing stakeholders with clear and accessible information about the decision-making process and the factors that are being considered.

Based on the requirements gathered in the co-design, the global architecture of the IDST is presented below.

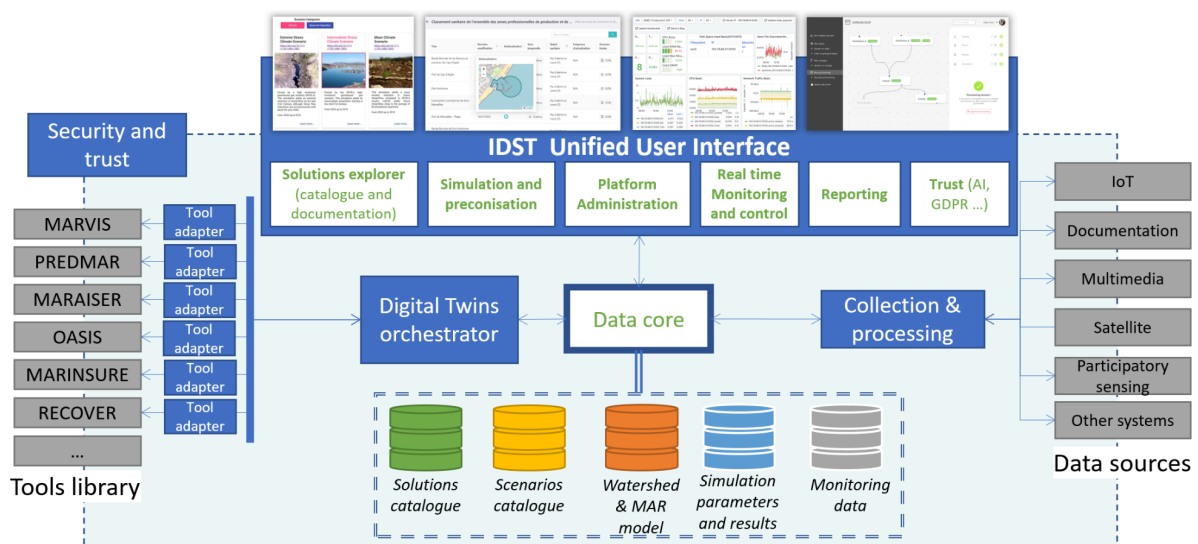


Figure 5: IDST global architecture

Its main components are the followings:

- IDST Data core
- IDST database catalogs and associated data models
- Data collection connectors
- Digital Twin orchestrator
- Tools management adapter
- User graphical interface

More detailed functionalities for each of the components are presented in the next sections.

4.1 IDST Data core

The IDST core component will be the **Stellio context broker**, an open-source implementation of the NGSI-LD specification.

In accordance with the NGSI-LD standard, it implements and exposes the three main families of APIs: the management of the information context, the temporal and geospatial search and the management of subscriptions and notifications.

Beyond the implementation of the NGSI-LD standard, it is fully part of a “*FIWARE architecture*”, both through the management of structural and contextualized data according to the common NGSI-LD data model, and through integration with the other bricks of the FIWARE ecosystem (*IoT agents, data collection, etc.*) made possible thanks to compliance with the NGSI-LD standard, guaranteeing the internal and external interoperability of the platform.

The principles implemented in the Stellio context manager are those of a modern, responsive and scalable architecture:

- Business microservices, divided according to the main NGSI-LD API families, subscribing to their topics of interest, and using databases adapted to their business.
- A pub-sub-type central message bus (Kafka) ensures the exchange of messages and events between platform components in a decoupled, extensible, scalable, and reactive manner.

The data produced by the various external components integrated into the platform (sensors, external data flow) are stored by the context manager in a PostgreSQL database with 2 specific extensions:

- Storing the *temporal evolution of the data* is done in the [TimescaleDB](#) database. Designed as an extension to PostgreSQL, it is specialized in the storage of temporal data (“*timeseries*”) and therefore perfectly suited, among other things, to the storage of sensor measurements, but more generally to any data that changes over time (for example, analysis performed at a specific point in time).
- Second, geo-information is managed with [PostGIS](#) (another extension to PostgreSQL) to allow spatial queries on the data stored in TimescaleDB (whether it is temporal or not).

4.2 Data models

Data models within the IDST will follow the NGSI-LD information model. The basic concepts of this model, also called the NGSI-LD Meta-Model, are **Entity**, **Relationship** and **Property**. The information model is derived from property graphs, with formal semantics defined on the basis of RDF and the Semantic Web framework. It can be serialized in the JSON-LD format. Each entity and relationship must be assigned a unique **IRI** (Internationalised

Resource Identifier), making the corresponding data exportable to the "Data in the Web". The suffix -LD refers precisely to this affiliation to the universe of Linked Data. The main components of a NGSI-LD data model are the followings:

- A NGSI-LD **entity** is the informational representation of something (a referent) that is supposed to exist in the real world, outside of the NGSI-LD platform. This referent does not need to be something strictly physical (it can be a legal or administrative entity) or stand-alone (it can be a system-level distributed construct). Any instance of such an entity is expected to be uniquely identified by a URI and characterized by a reference to one or more NGSI-LD entity types. In property graph language, it is a node.
- A NGSI-LD **property** is an instance that associates a characteristic, an NGSI-LD value, with an NGSI-LD entity, an NGSI-LD relationship, or another NGSI-LD property.
- A NGSI-LD **Relationship** is a directed link between a subject (starting point), which can be an NGSI-LD Entity, an NGSI-LD Property, or other NGSI-LD Relation, and an object (endpoint), which is an NGSI-LD entity.
- A NGSI-LD **value** is a JSON value (that is, a string, a number, true or false, an object, an array), or a value of type JSON-LD (that is, say a string as a lexical form of the value with a type, defined by an XSD base type or more generally an IRI), or a JSON-LD structured value (i.e. a set, a list, a labeled string in language).

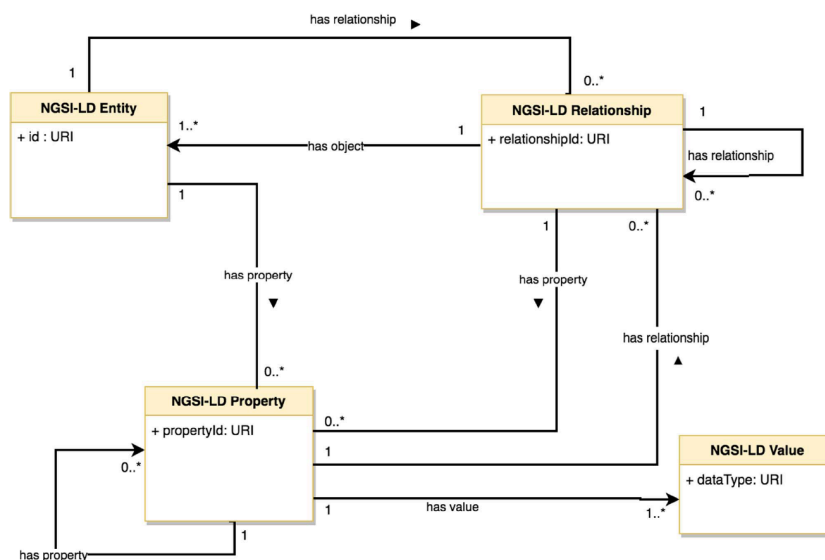


Figure 6: NGSI-LD information model

The actual data models definition will be heavily dependent on the specification of other architecture component definition, especially the tool management API, and will be provided in the next version of this deliverable.

4.3 Data collection connectors

The data collection will be performed via Apache NiFi which is an open source dataflow manager, leveraging the concept of ETL (extract, transform, load). NiFi is used to build a dataflow by assembling “processors”, it has a library of 400 processors that allows it to connect it to almost any data source to extract it, transform it and load it (to the IDST Data core).

The data collection and validation module provides tools and an environment for :

- The creation and visual management of a directed graph of processes, generally by composing a set of processes already existing in the catalog (which contains almost 400 ;
- A comprehensive catalog of existing processes, allowing native integration of a REST API (with or without authentication), processing of CSV, Excel, JSON or XML files, or access to files stored on FTP servers;
- Versioning of processing graphs, with traceability of modifications made to each graph, as well as the author of each modification;
- The ability to import and export treatment graphs;
- Automatic execution of processing flows at predefined intervals;
- The development of coherent, loosely coupled components that can simply be reused by different treatment processes;
- Error handling during processing, with the ability to configure specific actions when errors occur (retry, send notification, etc.);
- The ability, during processing, to query the database for duplicates and apply a specific treatment as required;
- Automatic, detailed tracking of all actions carried out during processing;
- Secure exchanges and the ability to define detailed authorizations;
- The ability to apply validations to processed data, notably via the definition of JSON or CSV schemas;
- Monitoring of all processing via a dashboard;
- Alerts triggered by missing or invalid data.

Moreover, this tool is natively designed for extensibility, and therefore offers all the functionalities required for the development and integration of specific processing components.

NiFi also provides an administration interface for processing chains, enabling customers to adapt them on a daily basis and add new file processing operations.

Historical datasets from the demo site have already been shared to all the consortium in the project shared folder. During the project, more data will be continuously collected via different means: existing and new to be installed IOT sensors, future lab analysis, ...

Some tools and components will work with continuous data flow to make some predictions of other variables, check in real time if contaminants don't breach a threshold to send some alerts, ... Hence, those new data will need to be transferred to the IDST data core in the most automated way possible, using the data collection connectors. Two types of data can be identified:

- **"Non automatic data"** -> sensor with manual reading, lab analysis results, ... -> A DATA TRANSFER folder was created in the project's Google Drive. Data are to be deposited here by the demo site, in the form of csv (to be preferred), Excel or PDF. Our expectation is that for each demo site, the format of the deposited files will be consistent throughout the project so that some NiFi flows can be created to automatically read any new file and inject its content to the IDST.
- **"Automatic data"** -> IOT sensors regularly send data automatically to any interface.
 - > For the PT demo site, we have deployed a FTP server in our infrastructure, the data from the spectrolyser (and possibly other automatic sensors), will be regularly sent to the FTP. Some processes will check the content of the FTP regularly and send the data to the IDST. The FTP server can be used by other demo sites if the same method fits their needs. The transfer method is still to be clarified with NE and SP demo sites.

4.4 Digital Twin Orchestrator and Tools management adapter

One of the most important parts of the MARCLAIMED IDST is to define how the tools will be configured, executed, displayed to the user. This specification will be built firstly based on the requirements of all 6 tools but will also be generic and flexible enough to allow future connection to other open-source tools as well as some ML models.

In order to keep this specification generic, it will describe the process to execute a "DataService" (the tools are themselves a kind of "Dataservice" that can be executed), referred to as simply services in the rest of the document, for simplicity.

Before being executed, each service needs to be configured with the relevant parameters that may vary between different tools. The first step in the service execution specification is thus to define how a service will be instantiated, i.e. an instance of a specific service with a defined set of configuration parameters is created, this instance can then be executed manually by the user or executed automatically by the IDST on a time based frequency or anytime a new input data is received (without the need to respecify the configuration parameters each time).

The sequence diagram for the creation of this service instance is presented hereafter. The different steps include the selection of the main configuration parameters such as the selection of the scenario (optimistic/pessimistic climatic projection, high/low population increase/decrease, ...) that is being used or the chosen asset, i.e. the infrastructure (which is linked to sensors, ...) + environment (which is linked to weather information, population, ...) on which the service will be executed.

This process results in the creation of a “DataServiceProcessing” entity, which corresponds to an instance of a service, that contains all the required information and configuration parameters to be able to run it without further user input.

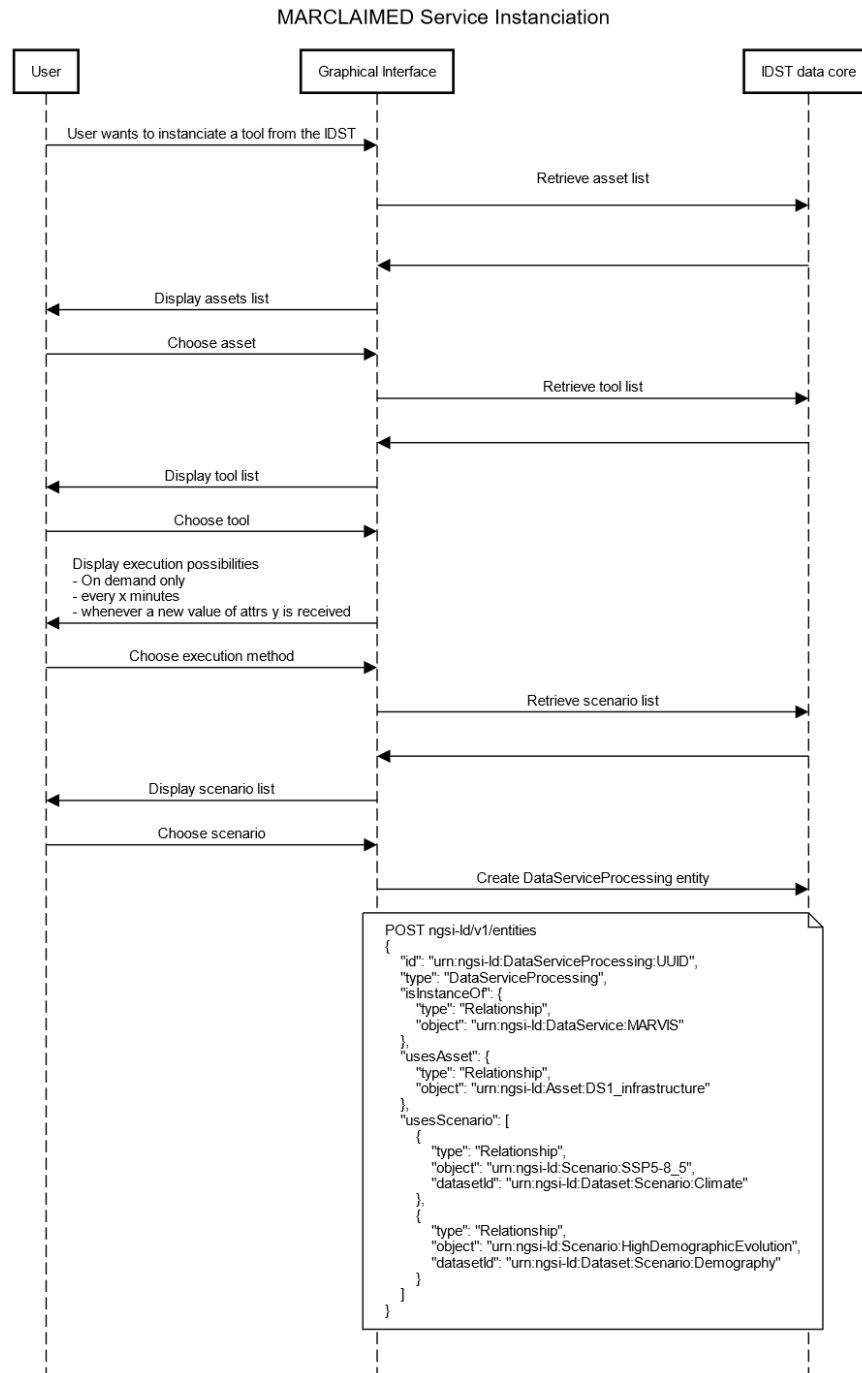


Figure 7: Service instantiation sequence diagram

Once a service is instantiated, it can be executed either manually by a user, or automatically. The sequence diagram for service execution manually launched by a user is presented hereafter.

Once the user asks for the execution, the IDST data core sends a notification to the Digital Twin Orchestrator, containing the content of the DataServiceProcessing entity, i.e., the configuration parameters. Based on those parameters, the orchestrator will gather all the required inputs from the data core and will send it all to the Tool management adapter. The tool management adapter will then convert the NGSI-LD entities that it received to whatever input format the service is expecting, then send it to the service, receive the output and convert the output to NGSI-LD, and send it in the chain, back to the data core to be then display to the user.

The Digital Twin Orchestrator is a unique process (there will be only 1 version of it) that will receive the DataServiceProcessing entities and retrieve all the required inputs from the data core, to send them to the relevant Tool management adapter.

For the Tool management adapter, there will be one version per service, that will make the specific conversion needed between the generic output of the Digital Twin Orchestrator and the specific input that each service requires.

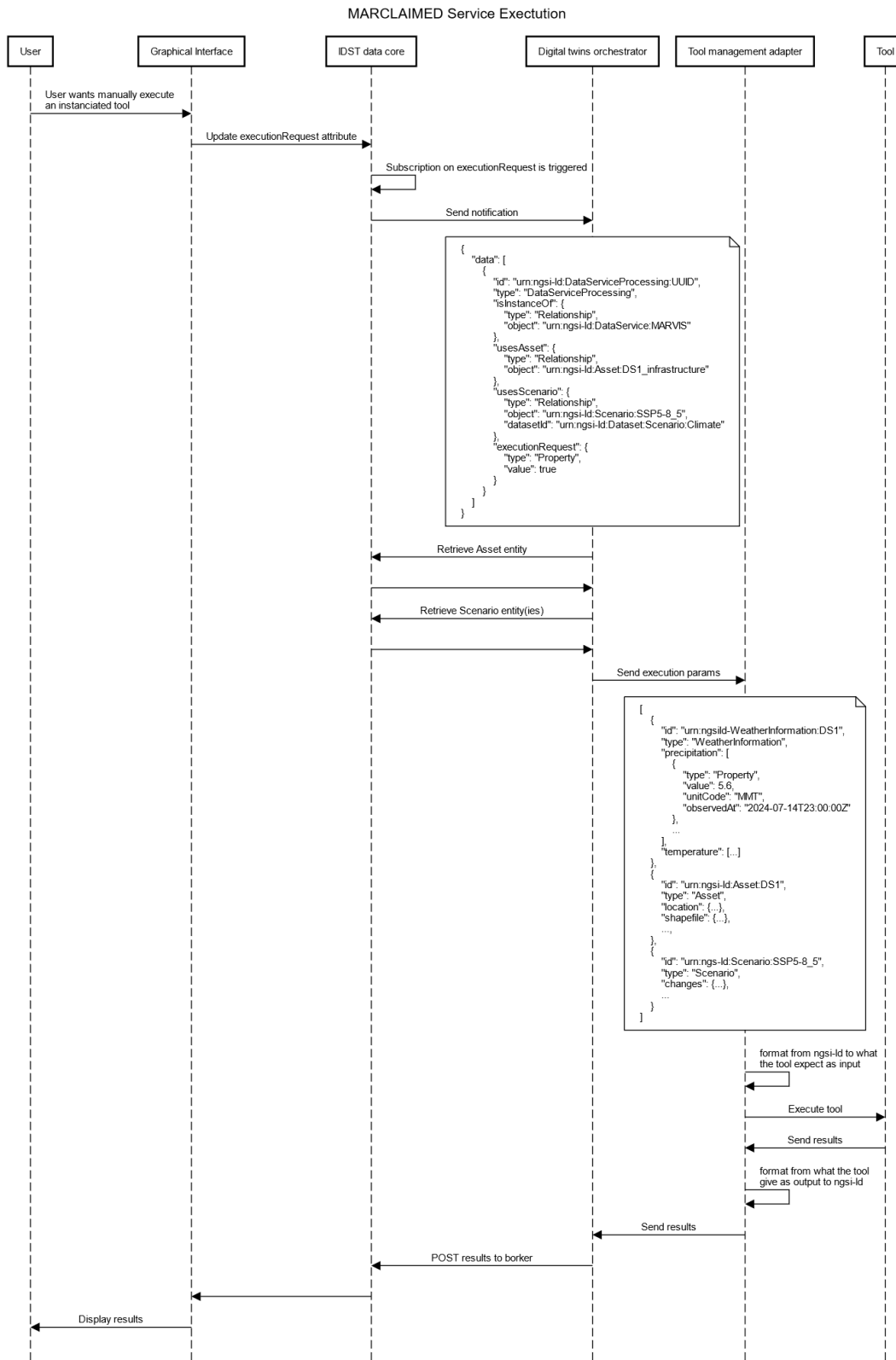


Figure 8: Service execution sequence diagram

4.5 IDST user interface

The IDST will feature a graphical user interface that will allow the user to:

- Explore the different available scenarios and MAR solutions stored in the database.
- Execute a simulation on a tool, visualize the results, compare them with results from other simulations.
- Visualize its historical data.
- Monitor its data in real time through dashboards.

The IDST will also be able to send some alerts and recommendations based on the results of some orchestrations.

Detailed specifications on the user interface will be provided later in the project, when the specifications for the tool execution is finalized.

4.6 Deployment

The IDST will consist of several software components that must interact with each other. These components must be installed on Linux-like servers. The whole thing must be secured to protect access to data and backups (replications) in place to prevent any data loss in the event of, for example, a hardware failure

The deployment of the IDST platform can be done in 2 main ways with regard to the different pilot sites:

1. As Software as a Service (SaaS): In this case, the platform is deployed in a cloud managed by EGM. This case is the simplest from a demo site point of view since it does not require any local server installation, since the access to the IDST is done through a web browser.
2. Deployed 'on premise'. This involves the provision and installation of Linux servers by the demo sites, including securing and replicating data. It is necessary in this case to have a suitable server room (with duplicated energy and network access). In addition, this makes as many IDST platforms to deploy and maintain.

The SaaS approach is therefore recommended in order to optimize resources and limit the environmental impact related to the deployment of several IT environments. It is important to emphasize that sharing a SaaS platform does not imply data sharing between demo sites:

- Stellio, the NGSI-LD context broker used as a centralized database, allows you to define 'tenants' whose data between Tenants is isolated as if they were several different databases. Each pilot site can thus manage its own data.

- Within each tenant, access controls on data entities in order to finely manage data rights.

When aiming for high availability of services, it is necessary to have redundancy at several levels. Depending on the high availability rate requested (for example, an availability rate of 99% ensures availability of approximately 361 days out of 365 while a rate of 99.5% ensures availability of more than 363 days out of 365), redundancies will have to be multiplied and, above all, their implementation speeds increased. Redundancy can be considered at several levels:

- Redundancy of services installed on different servers and failover from one server to another;
- Redundancy of server components and interconnection elements;
- Dynamic distribution of data across multiple hard drives (RAID, NAS, SAN, etc.);
- Storing backups in different geographical locations.

Redundancy at the service level will generally be based on a floating IP system (or IP failover) that will direct user requests to an operational server. If that server goes down, a standby server takes over, and user requests are directed to that new server. This implies that a synchronization of states between the initial server and the standby server are in place continuously.

Hardware redundancy is usually managed at the data center level. For example, in the case of RAID 5 storage, the data is scattered and replicated on a minimum of 3 hard drives. If a hard drive fails, the RAID system continues to work and the hard drive can be replaced by a technician. The new hard drive then joins the hard drive group without interrupting the data access service.

As far as backups are concerned, they imply longer recovery times in the event of a failure. This can range from several minutes to several hours depending on the volume stored, the type of storage and its geographical location.

A backup is ideally based on a 3-2-1 system that allows you to increase the number of backups kept and diversify the places where they are stored. The rule states that at least three copies of the data must be made. Two of the backups are stored on different types of media, and at least one backup must be stored offsite or in the cloud. The 3-2-1 backup rule is a simple strategy that is easy to remember and follow. This is a reliable method that businesses can follow.

To present these approaches, we speak of a business continuity plan (BCP) and a disaster recovery plan (DRP)

- The BCP aims to guarantee the high availability of the computer system, particularly in the event of a crisis. It is nothing more and nothing less than ensuring that all the critical applications necessary for the company's activity are available, even in the event of a disaster. It relies on system redundancy and can become very expensive when high levels of availability are required
- The DRP is there to manage the risk of service interruption. If the information system is no longer available (hardware failure, cyberattack, etc.), the DRP will describe all the procedures necessary to restart the computer system as quickly as possible. This restart cannot be done in any way, and the DRP is there to define the nature and order of the actions to be taken to restore the system and data to the state they were in before the incident. It relies on backups. In particular, a DRP analyzes the following points:
 - o the identification of critical activities that absolutely must be continued,
 - o the identification of all risks,
 - o the different maintenance solutions,
 - o the procedures to be applied according to different scenarios,
 - o the resources necessary for business continuity (people, raw materials, equipment, subcontractors, etc.),
 - o the recovery deadlines to be respected.

For example, for IDST, monitoring and warning systems will be more critical than scenario evaluation systems.

These different points will be taken into account when deploying the servers and a DRP will have to be written and integrated into D2.2.

At the very least, a database clustering service should be considered, as a basis for synchronizing states between servers. The envisioned schema is presented in the next figure.

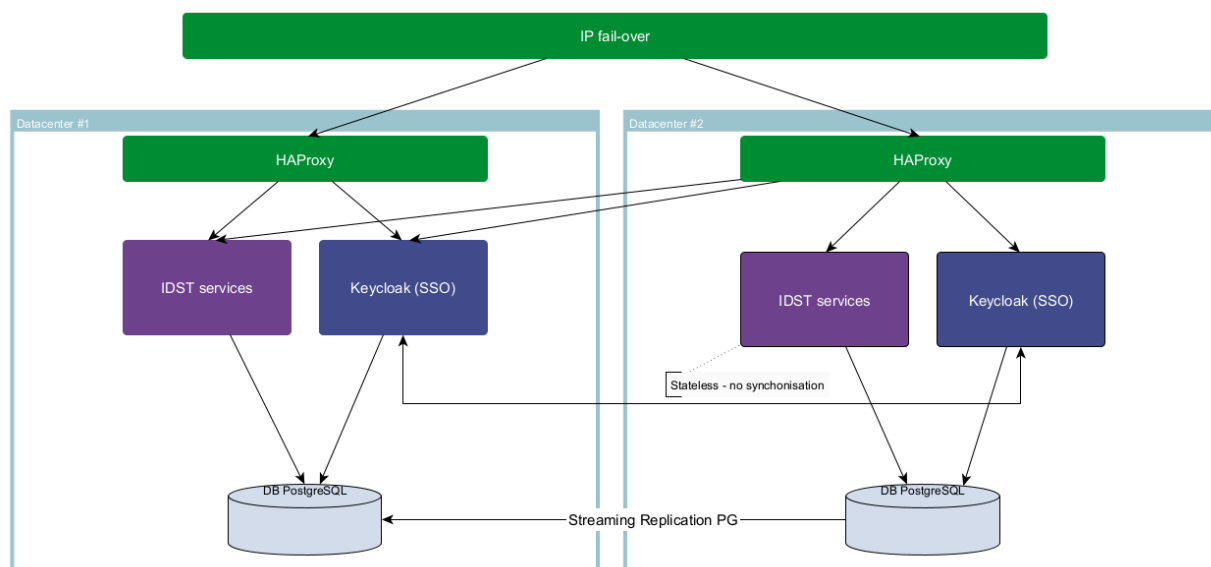


Figure 9: Envisioned distributed architecture

5 CONCLUSIONS

This deliverable provides the first versions of the requirements gathered to design the MARCLAIMED IDST. Each demo site is unique, with distinct local specificity and a large variety of stakeholders, each having their own needs and expectations regarding MARCLAIMED and the IDST. In addition, the IDST will integrate 6 different tools involving different kinds of input, triggering mode, deployment, etc.. Therefore, the IDST is starting to specify the module that will instantiate and execute those tools, through the Digital Twin Orchestrator and the Tool management API. In addition, the other core modules of the IDST, such as the data core, the import/export data connectors, and the graphical interface are described in this first version of a global architecture. Finally, solutions regarding the future ways to deploy all the modules were presented and compared.

This architecture will be continuously refined throughout the project, and will be deployed early in the project as part of task T2.2, allowing for actual testing leading to further improvement.

REFERENCES

- Angelakis, A. N., Marecos do Monte, M. H., Bontoux, L., & Asano, T. (1999). The status of wastewater reuse practice in the Mediterranean basin: Need for guidelines. *Water Research*, 33(10), 2201-2217.
- Bixio, D., Thoeye, C., De Koning, J., Joksimovic, D., Savic, D., Wintgens, T., & Melin, T. (2005). Wastewater reuse in Europe. *Desalination*, 187(1-3), 89-101.
- Bouwer, H. (2002). Artificial recharge of groundwater: hydrogeology and engineering. *Hydrogeology Journal*, 10(1), 121-142.
- Dillon, P. (2005). Future management of aquifer recharge. *Hydrogeology Journal*, 13(1), 313-316.
- Dolnicar, S., & Schäfer, A. I. (2009). Desalinated versus recycled water: public perceptions and profiles of the accepters. *Journal of Environmental Management*, 90(2), 888-900.
- Foster, S., Cilton, J., Moench, M., Cardy, F., & Schiffler, M. (2002). Groundwater in rural development: facing the challenges of supply and resource sustainability. The World Bank.
- Gale, I. (2005). Strategies for Managed Aquifer Recharge (MAR) in semi-arid areas. UNESCO-IHP, Paris.
- Herman, J., Zeff, H. B., Reed, P. M., & Characklis, G. W. (2012). Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty. *Water Resources Research*, 50(10), 7692-7713.
- Lazarova, V., & Bahri, A. (2005). Water reuse for irrigation: Agriculture, landscapes, and turf grass. CRC Press.
- NRMMC-EPHC-NHMRC. (2009). National Water Quality Management Strategy: Australian Guidelines for Water Recycling. National Resource Management Ministerial Council, Environment Protection and Heritage Council, National Health and Medical Research Council.
- Pahl-Wostl, C., Jeffrey, P., Isendahl, N., & Brugnach, M. (2011). Maturing the new water management paradigm: progressing from aspiration to practice. *Water Resources Management*, 25(3), 837-856.
- Po, M., Kaercher, J. D., & Nancarrow, B. E. (2003). Literature review of factors influencing public perceptions of water reuse. CSIRO Land and Water.
- Pyne, R. D. G. (2005). Aquifer storage recovery: A guide to groundwater recharge through wells. ASR Press.
- Rauch, W., Seggelke, K., Brown, R., & Krebs, P. (2005). Integrated approaches in urban stormwater management. *Water Science and Technology*, 52(9), 165-173.

Scanlon, B. R., Reedy, R. C., Stonestrom, D. A., Prudic, D. E., & Dennehy, K. F. (2006). Impact of land use and land cover change on groundwater recharge and quality in the southwestern USA. *Global Change Biology*, 11(10), 1577-1593.

Schnoor, J. L. (2010). Water sustainability in a changing world. *Environmental Science & Technology*, 44(13), 4849-4850.

Vörösmarty, C. J., Green, P., Salisbury, J., & Lammers, R. B. (2000). Global water resources: vulnerability from climate change and population growth. *Science*, 289(5477), 284-288.

MARCLAIMED will support decision makers to integrate Managed Aquifer Recharge (MAR) with Alternative Water Resources (AWR) in River Basin and Drought Management Plans supporting the adaptation and resilience of structural supply systems in the context of climate change. The main outcome will be the MARCLAIMED Integrated Decision Support Tool (IDST) that will provide solutions to tackle existing and emerging threats and support decision making and adaptation policies.



MARCLAIMED



**Funded by
the European Union**