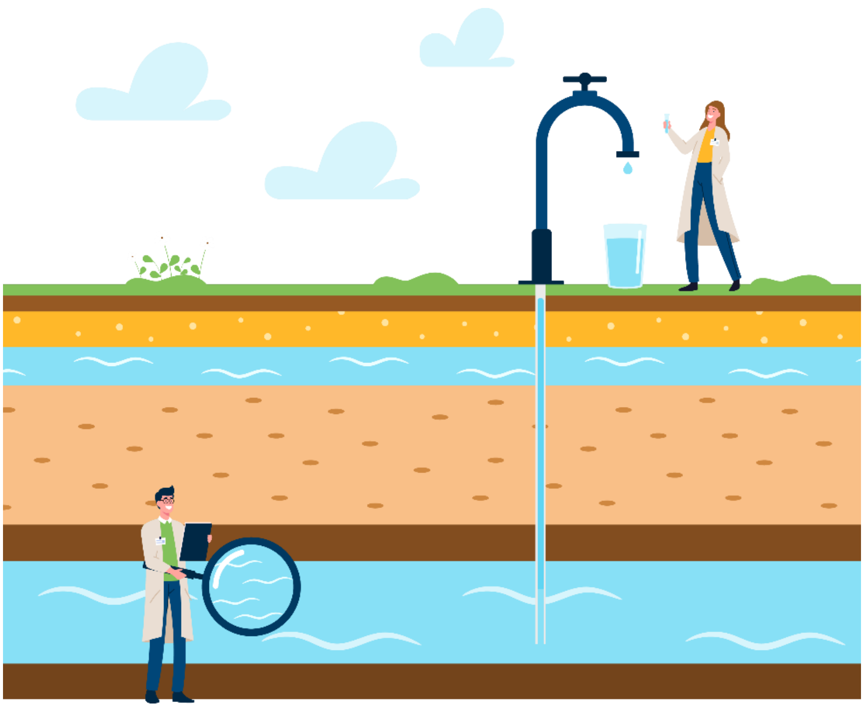
D4.2 Report of M-AI-R DSS data architecture and data space technical specifications

CETAQUA ANDALUCÍA



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101082048

Funded by the European Union

GRANT AGREEMENT NUMBER: 101082048

PROJECT ACRONYM: MAR2PROTECT

PROJECT TITLE: "Preventing Groundwater Contamination Related to Global and Climate Change through a Holistic Approach Based on Managed Aquifer Recharge"

PROJECT Duration: 1st December 2022 - 30th November 2026 (48 months)

WEBSITE: https://mar2protect.eu/

ABBREVIATION / ACRONYM:

Abbreviation / Acronym	Description
API	Application Programming Interface
CSV	Comma-Separated Values
DS	Demo Site
DSS	Decision Support System
GIS	Geographic Information System
GW	Groundwater
IoT	Internet of Things
MAR	Managed Aquifer Recharge

Executive Summary

The following document is Deliverable 4.2 "Report of M-AI-R DSS data architecture and data space technical specifications" of the MAR2PROTECT project, funded by the European Union's Horizon Europe research and innovation programme under grant agreement Number 101082048.

This document aims to define the architecture that the M-AI-R DSS will follow.

Deliverable Number	WP / T
D4.2	WP4.T4.1
Lead Beneficiary	**Deliverable Author (S)**
Cetaqua Andalucía	Ángel Cañete (Cetaqua Andalucía) Sara Espinosa (Cetaqua Andalucía) Manuel Argamasilla Ruiz (Cetaqua Andalucía)
Beneficiaries	**Deliverable Co-Author (S)**
Organisation	Andrew Watson (SUWI) Catarina Novo (IT) Ming Cheng (UNIBO)
Planned Delivery Date	**Actual Delivery Date**
30/11/2023	28/11/2023

Type of deliverable	PU	Public, fully open, e.g. web (Deliverables flagged as the public will be automatically published in the CORDIS project's page))	X
	SEN	Sensitive, limited under the conditions of the Grant Agreement	

ACKNOWLEDGEMENT: The work described in this report has been funded by the European Union from the Horizon Europe Research and Innovation Programme (HORIZON-CL6-2022-ZEROPOLLUTION-01), Research and Innovation Action under the Grant Agreement No 101082048.

Index

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101082048
Funded by the European Union

1. Introduction

The M-AI-R DSS is a tool to efficiently manage groundwater recharge management using both conventional and non-conventional resources, from design to operation. In the design phase, the suitability and viability of the recharge will be taken into account, and in the operation phase, the tool will contribute to the management of the risk associated with the operation itself (infrastructure and state of the groundwater).

This document describes the different tools and platforms that make up the M-AI-R DSS and explains how these tools communicate between them for its effective operation.

Data architecture is the organized framework that dictates how data is structured, stored, processed, and monitored within a software system. It serves as the guiding plan that directs the establishment and maintenance of the data-related elements of the system, ensuring that data is transformed into a valuable asset. This document will comprehensively delineate the data architecture that enables efficient information exchange among the tools of the M-AI-R DSS platform. Our aim is to provide a comprehensive understanding of how data is structured, stored, processed, and managed within the context of the project. By delving into the intricacies of data architecture, this deliverable will establish a solid foundation upon which the project can effectively harness its data resources to drive innovation, optimize performance, enhance security, ensure compliance, and ultimately achieve its strategic objectives.

The components of the M-AI-R-DSS are the following:

M-AI-R DSS: It is the core of the system, and its objective is to facilitate the development of groundwater management strategies. The DSS aims to empower decision-makers by enabling them to: i) identify appropriate locations for Managed Aquifer Recharge (MAR) projects; ii) design MAR projects in a cost-effective manner; iii) develop risk assessment and mitigation measures for MAR, taking into account global climate change and climate change scenarios; and iv) provide the information to the visualization platform.

Visualization platform: Using the information provided by the M-AI-R DSS, the visualization platform shows the results of all MAR-related activities through a user-friendly interface.

RAINREC: The RAINREC tool is a model used to accurately evaluate the impact of climate change on aquifer recharge. It focuses on redistributing rainfall-runoff recharge and can be applied in various scenarios.

DRONE: This tool is an innovative, open-source modelling tool designed to assess the effects of climate change on groundwater. It utilizes hybrid ARMA/ARDL models and nonstationary time series to predict how aquifers will respond to long-term climate change.

REACH: A digital tool for improving the understanding of the main processes conditioning groundwater (GW) chemical status is being developed. This innovative tool combines advanced data analytics and modelling techniques to provide reliable

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101082048
Funded by the European Union

forecasts of the risk of GW pollution, enabling stakeholders to make informed decisions and implement effective measures to protect GW quality.

SUIT & FEAS: Within the Decision Support System (DSS), there will be a specific tool that evaluates the spatial suitability and practicality of various MAR options. This module aims to create GIS maps indicating the suitability of MAR and identify potential areas of interest or MAR projects.

GWPrev Platform: The GW-PREV IoT platform controls the MAR progress using information provided by real-time integrated sensing systems and innovative analytical techniques for the monitoring of pollutants, generating a better understanding of pollution sources and pathways in groundwater. This platform will have one section per demo site.

It is important to note that the tools encompassed within the data architecture are currently in various stages of development. Consequently, some of the details and information contained within the data architecture may undergo revisions and adjustments as these tools continue to evolve.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101082048

Funded by the European Union

2. M-AI-R DSS Architecture and Data Space

This section delves into the underlying structure and organization of the M-AI-R- DSS platform, highlighting the key components and their interconnections.

2.1. General overview

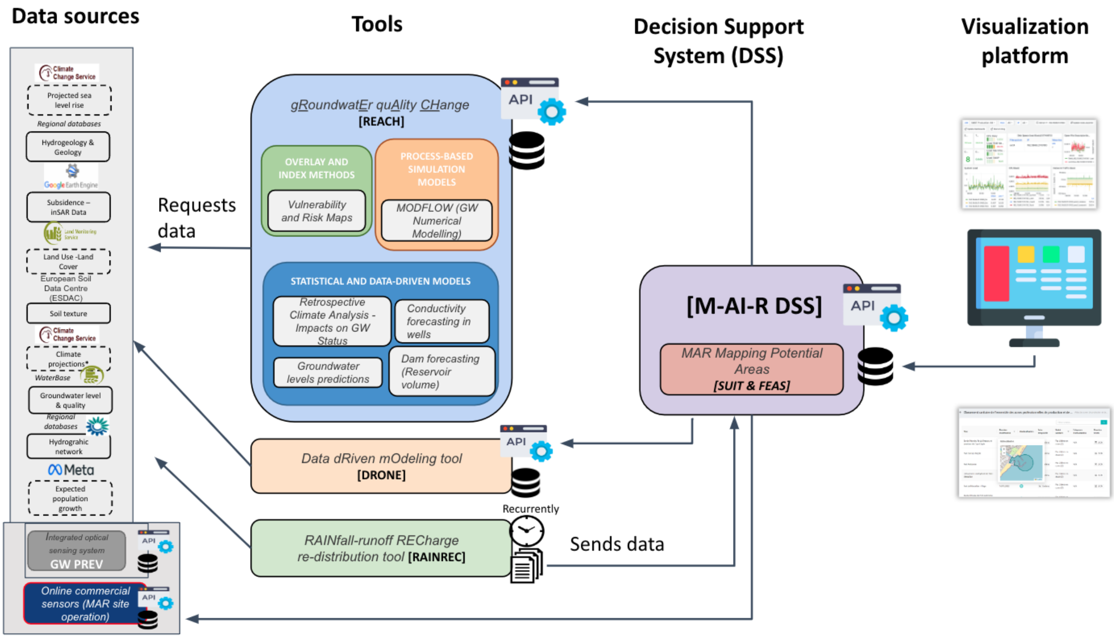


Figure 1: Full-view of the proposed platform's architecture

Figure 1 presents a general overview of the proposed architecture for the M-AI-R DSS platform. The platform consists of seven main components: GW PREV, RAINREC, DRONE, REACH, SUIT & FEAS, the M-AI-R DSS which is the core of the system, and a visualization platform. The GW PREV platform, RAINREC, DRONE, REACH, and SUIT & FEAS tools, that serve information to the M-AI-R DSS, are independent but complementary and retrieve essential data from open-source repositories, as well as the IoT (Internet of Things) sensors deployed at the different demo sites where these modules will be operating. Finally, the information provided by the M-AI-R DSS is accessible through a visualization platform.

The tools integrated within the platform draw upon a diverse array of data sources, encompassing both open-source repositories and information retrieved from IoT sensors strategically deployed at distinct demonstration sites. These sources of data serve as the lifeblood of the tools, providing essential inputs that enable them to operate effectively and deliver insightful outcomes. Moreover, each tool is accompanied by a dedicated database, diligently maintaining a comprehensive record of the outputs they generate at regular intervals. This database functionality is instrumental in preserving historical data, thereby facilitating further analysis, comparisons, and decision-making processes, which are vital components of our comprehensive data architecture strategy.

At specific time intervals, the M-AI-R DSS initiates data requests from the components to ensure the currency of information within its linked database. The periodicity of these updates varies among the tools, aligning with the frequency at which they, in turn, acquire and provide information. This frequency is contingent upon the specific data sources employed by each tool for generating their respective outputs. As shown

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101082048
Funded by the European Union

in Figure 1, the GW PREV, DRONE, REACH, SUIT & FEAS, and M-AI-R DSS use Application Programming Interfaces to share the information contained in their databases with the rest of the components. An Application Programming Interface, commonly known as an API, serves as the bridge between different software applications, enabling them to communicate, interact, and share data seamlessly. APIs define a set of rules, protocols, and tools that allow developers to access the features and data of one software system from another. This architecture has been thoughtfully designed to foster seamless data exchange among the different tools that form the platform. The adoption of a modular approach and the integration of these APIs hold the potential to greatly augment our system's versatility, scalability, and interoperability. The establishment of a robust API infrastructure forms a cornerstone for future development and integration initiatives, allowing us to adapt and extend our architecture as our tools progress through their developmental phases and as changes in data sources and requirements may occur. In the case of the RAINREC tool, it will send the results of its executions to the M-AI-R DSS by means of files. In this regard, an endpoint will be enabled in the API of the M-AI-R DSS to allow the RAINREC tool to share its results. Once received, the M-AI-R DSS will automatically process the data and update the result in its database, which can then be consulted at any time from the visualization platform.

Regarding the specification of these APIs, it is worth emphasizing that the definition of their concrete functionalities will be meticulously delineated during the tools' implementation phase. This meticulous API specification process will yield a clear and comprehensive set of interfaces that encapsulate the functionality and data exchanges between the various tools. This comprehensive specification will play a pivotal role in guiding the work of developers engaged in the implementation of individual tools, ensuring a harmonious and interoperable ecosystem.

2.2. Data sources

As previously mentioned, each tool comprising the M-AI-R DSS platform functions as a self-contained entity, sourcing the necessary data from open-source data repositories, IoT sensors deployed in the respective demonstration sites, and, when applicable, other tools. These open data sources vary in their data retrieval methods, such as downloadable CSV/Excel files that require subsequent processing. Our primary efforts thus far have revolved around identifying and streamlining data acquisition from these sources, automating the process wherever feasible. For instance, in the case of demo-site 6, the M-AI-R-DSS tool leverages accumulated precipitation data from the hydrological year and precipitation forecasts to proactively assess potential aquifer recharge scenarios.

Concerning the IoT sensors, the GW PREV tool assumes the responsibility of furnishing this data to the other tools via an API. The REACH, DRONE, SUIT & FEAS, RAINREC, and M-AI-R DSS tools will make use of this data before it is ultimately utilized by the visualization platform.

In MAR2PROTECT, the M-AI-R-DSS will be implemented in at least 5 of the 7 demo sites (DS). The project's demo sites are illustrated on the MAR2PROTECT website (https://mar2protect.eu/), under the tab "demos".

The following tables summarize the data sources that are expected to be used in demo sites 2-6. Specifically, for each data source, the following details are given: the type of data, its name/acronym, how long the data is collected, the frequency of updates, and the type of access. Access on-demand involves the manual retrieval of information by requesting it from the technician, while public access data sources offer the flexibility of being accessible at any time via a URL. Such data sources may be downloadable, with the website providing a file that can be processed, or non-downloadable, where the information is directly available on the website. Cells containing a hyphen (-) are unknown data.

Table 1: Data sources and collected data for DS2 (Oued Souhil, Nabeul, Tunisia)

Kind of data	Data source	Data length	Data frequency	Access
Meteorological	MAWRF	17/11/2015, 28/11/2022	Hourly	Public access downloadable
	NIM	-, present	-	On-demand
GIS	-	N/A	-	On-demand
Hydrodynamic	DGRE	2020, - (requested every 5 years)	Twice a year	On-demand
Hydrochemical	DGRE	2020, -	Twice a year	On-demand

Table 2: Data sources and collected data for DS3 (Frielas, Portugal)

Kind of data	Data source	Data length	Data frequency	Access
Meteorological	SNIRH	2001, -	Hourly	Public Access downloadable
GIS	SNIRH	N/A	Monthly	Public Access
Hydrodynamic	SNIRH	1976, -	Monthly	Public Access
Hydrochemical	SNIRH	1976, -	Every 6 months	Public Access
Meteorological	SNIRH	2001, -	Hourly	Public Access downloadable

Table 3: Data sources and collected data for DS4 (Emilia Romagna, Italy)

Kind of data	Data source	Data length	Data frequency	Access
Meteorological	ARPAE (Rainfall, Temperature, Evapotranspiration, land vertical)	1916, -	Daily	Public Access downloadable
	SONEL, PSMSL (Sea water level)	-	-	Public Access
GIS	-	N/A	-	-
Hydrodynamic	ARPA	2018, -	Monthly	Private Access

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101082048

Funded by the European Union

Hydrochemical	-	2018, -	Monthly	Private Access

Table 4: Data sources and collected data for DS5 (Cape Flats, South Africa)

Kind of data	Data source	Data length	Data frequency	Access
Meteorological:	Hortec	2018-2023	Daily	Private Access
	NOAA, DWS, SASSCAL	1970-2019	Daily	Public Access downloadable
	SAWS, ARC	1970-2019	Daily	Private Access
	CORDEX	1970-2050	Daily	Public Access downloadable
GIS	Elevation	N/A		Public Access downloadable
	Soils	N/A	N/A	Public Access downloadable
	Geology	N/A	N/A	Private Access
	SANLC (landcover)	2018	-	Public Access downloadable
Hydrodynamic	ARPA	2018, -	Monthly	Private Access
Isotopes	IsoRSM/IsoGSM (Precip isotopes)	1970-2023	Daily	On-demand
	GNIP, GNIR (IAEA)	1970-2023	Daily	Public Access downloadable
	Student collection	2022-2024	Daily	On-demand
Hydrology	DWS	1970-2023	Daily	Public Access downloadable
Reservoir operation	Stellenbosch Municipality			
Hydrochemical	Student collection	2022-2024	Daily	On-demand
	National Groundwater Archive	1988-1995/2023-2025	Daily	Public Access

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101082048

Funded by the European Union

Table 5: Data sources and collected data for DS6 (Marbella, Spain)

Kind of data	Data source	Data length	Data frequency	Access
Meteorological:	SAIH	1994, -	Hourly	Public Access downloadable
	SAIH (rainfall forecasting)	N/A	Daily	Public Access downloadable
GIS	IGME, IGN, DERA, REDIAM	N/A	-	Public Access
Hydrodynamic	HIDRALIA	2000, present	Monthly	On-demand
Hydrochemical	National Groundwater Archive	2011, -	Irregular (control point)	Public Access downloadable

2.3. M-AI-R DSS components

Figure 2 shows the classification of the expected results depending on the technology applied, as well as the DS in which the tools are expected to be implemented (the visualization platform does not appear for not being a logical component).

The type of results identified are:

- **Descriptive:** What is happening? This type of result is provided by the GW PREV IoT platform, RAINREC Tool, the Numerical modelling module of the REACH Tool and Vulnerability and Risk Maps of the REACH Tool.
- **Diagnostic:** Why is it happening? These results are provided by the Retrospective Climate Analysis module of the REACH Tool and also by the RAINREC Tool and the Numerical Modelling module of the REACH Tool.
- **Predictive:** What is going to happen? This result is provided by the DRONE Tool, Groundwater levels prediction module of the REACH Tool, RAINREC Tool and the Numerical Modelling module of the REACH Tool.
- **Descriptive:** What is the optimal course of action or recommendation? Provided by M-AI-R DSS.

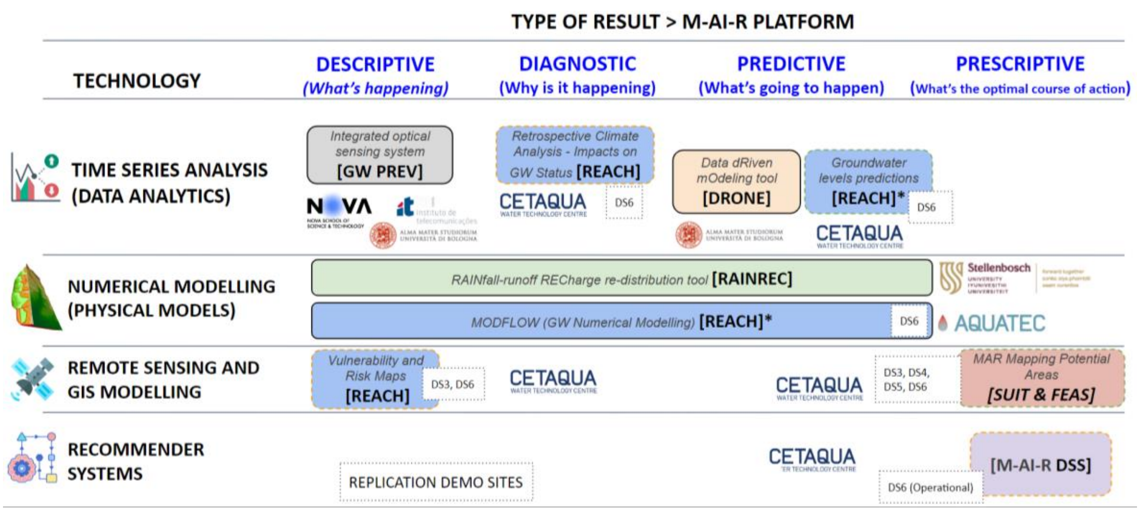


Figure 2: Classification of expected results and Demo Sites (DS) replication

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101082048
Funded by the European Union

Also, a preliminary identification of the DS replication of each tool has been done:

- The Vulnerability and Risk Maps module of the REACH Tool will be implemented in DS3.
- The SUIT & FEAS tool will be implemented in DS3, DS4, DS5, and DS6.
- M-AI-R DSS will be implemented in DS6.

2.3.1. Inputs and outputs of the components

DRONE: This tool uses data inputs on electrical conductivity, which serves as a proxy for salinity, as well as temperature. These parameters are essential for the final model, and it is certain that they will be incorporated into its final version. Additionally, other indicators may be incorporated to gather supplementary information that may provide further insights, as the Drone tool is under deployment. As a result, the tool returns the forecasting in electrical conductivity and temperature, with a forecast space to be defined according to project requirements and results reliability (the larger the forecast space, the lower the reliability).

RAINREC: This tool uses data inputs on temperature, wind, rainfall, relative humidity, solar radiation, streamflow, isotopes precipitation, isotopes streamflow, isotopes groundwater, groundwater levels, reservoir release, Sewage works outflow for current climate conditions. Climate change projections will then be used as future scenarios for the above-mentioned data. As an output, the RAINREC tool returns predictions on electrical conductivity, surface runoff, interflow, baseflow, streamflow, soil moisture, isotopes surface runoff, isotopes interflow, isotopes baseflow, isotopes streamflow, isotopes soil moisture and groundwater levels for current and future scenarios. Finally, the outputs will be fed into a data-driven approach to determine the main determinants of groundwater total dissolved solids (as a bulk indicator of water quality) to generate maps of current and future groundwater total dissolved solids with and without MAR.

REACH: This tool uses data inputs on the volume of water released, rainfall, temperature, evapotranspiration, and expected precipitation to predict the volume of water stored. On the other hand, the REACH tool is expected to use historical data on the piezometric level of the aquifer and water withdrawals to predict the piezometric level of the aquifer. Finally, depending on the amount and quality of data, it is possible that the REACH tool uses historical data on the electrical conductivity of the aquifer for its prediction. The forecast space of the aforementioned predictions has not been defined yet.

SUIT & FEAS: This tool uses as an input geographical data (GIS) and returns as output the representation of areas where it is feasible and advisable to carry out artificial recharge of aquifers, based on previous maps of vulnerability to contamination in aquifers.

GWPrev Platform: Most of the planned sensors will be optical-based. The input will be the optical interaction between the measurand, and the fluorophore compound and the output of the sensor will be the concentration of the measurand (compound of interest).

M-AI-R DSS: This tool, which is the core of the platform, will use the output of the rest of the tools and platforms to facilitate the development of groundwater management strategies. It also collects and provides the information to be displayed on the visualization platform.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101082048

Funded by the European Union

Visualization platform: It only demands data from the M-AI-R DSS, which collects the information to be displayed within the platform.

2.4. Data structure

The development of the APIs within the M-AI-R DSS platform is presently in its construction phase. Given their ongoing evolution, it is regrettably impractical to furnish a comprehensive definition of these APIs within the scope of this document. These APIs represent a pivotal element of our dynamic data architecture, and as they progress toward a more mature and stable state, their precise functionalities and specifications will be clearly delineated. Nevertheless, while detailed API descriptions are pending, this document will certainly offer essential insights into the nature of data files shared between modules.

The data files exchanged between modules will be formatted in CSV (Comma-Separated Values) format. This choice of CSV format brings significant advantages to our data processing workflows. CSV is a widely recognized and platform-agnostic format, ensuring compatibility across various software systems and environments. The semicolon as a delimiter provides versatility, especially when dealing with data that may contain commas. Furthermore, CSV files are human-readable, making it easy to inspect and troubleshoot the data. Their compact structure reduces storage space requirements and minimizes data transfer times, which is particularly advantageous when working with large datasets. Additionally, the CSV format simplifies data import and export procedures, streamlining the exchange of information between different modules and enhancing the overall efficiency and reliability of our data processing operations.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101082048

Funded by the European Union

3. Conclusions

This document describes the data architecture that the tools that make up the M-AI-R DSS will follow for their correct operation. The data architecture serves as a framework that defines how data will be collected, stored, processed, and analyzed within the M-AI-R DSS ecosystem. It encompasses the design principles, data models, and data integration strategies that will be employed to ensure the efficiency, reliability, and accuracy of the system.

The data architecture presented in this document is based on the current understanding and analysis of the M-AI-R DSS requirements. It outlines the various components, databases, and interfaces that will be utilized to support the functionalities of the system. It provides guidelines for data governance, data quality assurance, and data lifecycle management to ensure that the right data is available at the right time and in the right format for decision-making processes.

It is important to note that these tools are still under development, and as such, the data architecture presented herein is subject to further refinements and enhancements. The iterative nature of software development means that new requirements may surface throughout the development lifecycle. Therefore, it is crucial to remain adaptable and open to incorporating new data architecture requirements that may arise during the development process.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101082048

Funded by the European Union