D1.2 – Self-Assessment and Data Management Plan

WP1 Project Management

Author: Gerasimos Antzoulatos (CERTH), Ilias Gialampoukidis (CERTH), Agni Konstantakou (CERTH), Stefanos Vrochidis (CERTH)

Date: 30-04-2023

Co-funded by the European Union

GRANT AGREEMENT NUMBER	101070262		
FULL TITLE / ACRONYM	Water Data Management Ecosystem for Water Data Spaces / WATERVERSE		
START DATE	1 October 2022	DURATION	36 months
END DATE	30 September 2025		
PROJECT URL	https://waterverse.eu/		
DELIVERABLE	D1.2 Self-assessment and data management plan		
WORK PACKAGE	WP1		
CONTRACTUAL DATE OF DELIVERY	30 April 2023		
ACTUAL DATE OF DELIVERY	02 May 2023		
TYPE	DMP — Data Management Plan	DISSEMINATION LEVEL	Public
LEAD BENEFICIARY	CERTH		
RESPONSIBLE AUTHOR	Gerasimos Antzoulatos (CERTH), Ilias Gialampoukidis (CERTH), Agni Konstantakou (CERTH), Stefanos Vrochidis (CERTH)		
CONTRIBUTIONS FROM	CET, KWR, ENG, PHOEBE, EGM, WE, FIWARE, PWN, HIDR, KEY, WBL, HST, UNEXE, SWW		
ABSTRACT	The document outlines the assessing plan of the project objectives, with quantitative measures and indicators where appropriate, summarised in form of tables. The deliverable also describes the data management plan that will be constantly updated, and the final version will be delivered at M20.		

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.

Copyright message

REVISION HISTORY

Version	Date	Who	Description
0.1	13-02-2023	CERTH	First release of the template, Table of Contents
0.2	27-02-2023	CERTH	Compose self-assessment plan (section 2)
0.3	10-03-2023	CERTH	Compose Data Management Structure (Section 3)
0.4	17-03-2023	CERTH	Compose DMP for WATERVERSE Datasets (Section 4)
0.5	31-03-2023	CERTH	Release new version based on the contribution from involved partners CET, KWR, ENG, PHOEBE, EGM, WE, FIWARE
0.6	10-04-2023	CERTH	Release new version based on the contribution from involved partners PWN, HIDR, KEY, WBL, HST, UNEXE, SWW
0.7	13-04-2023	CERTH	Release version for internal reviewing
0.8	29-04-2023	CERTH	Addressed comments of reviewers
1.0	30-04-2023	CERTH	Final version for submission

QUALITY CONTROL

Role	Date	Who	Approved/Comment
Internal review	25-04-2023	CET	Approved with some minor revisions and comments
Internal review	25-04-2023	FIWARE	Approved with some minor revisions and comments

EXECUTIVE SUMMARY

In this deliverable, the first version of the WATERVERSE’s Self-Assessment Plan (SAP) is presented that assesses the degree of fulfilment of the project’s objectives. The aim is to provide a methodology on how the quality of the scientific and research achievements that will be conducted in this project’s context, is being successfully monitored and assessed periodically, covering throughout the project’s lifetime. The SAP has been broken down by relying on the project's Scientific Objectives (SOs) that are measured via specific Key Performance Indicators (KPIs).

Furthermore, the first version of WATERVERSE’s Data Management Plan (DMP) is introduced providing a general outline of the project policy concerning the data management. The proposed policy reflects the agreement among the consortium regarding the activities that should be taken for data management and that are consistent with EU regulations. The DMP has the objective to detail specifics of data that have already been collected or generated during the lifespan of the project. Also, it contains details on how the datasets will be properly handled, documented, shared and stored. Moreover, it includes a summary of the data and how they will be FAIR (i.e., Findable, Accessible, Interoperable, and Reusable). In general, the overall purpose of the DMP is to support the data management life cycle for all data that will be collected, processed, or generated by the project. It is a living document that will be updated periodically during the project’s lifetime. The datasets may also be altered due to various reasons, such as project maturity, legislative changes, etc. This document will evolve along the project, and it will be updated in D1.3 Self-assessment and data management plan v2 (M20).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

CA	Consortium Agreement
DMP	Data Management Plan
DoA	Description of Actions
DOI	Digital Object Identifier
FAIR	Findability, Accessibility, Interoperability, and Reuse
GA	Grand Agreement
GDPR	General Data Protection Regulation
KPI	Key Performance Indicator
ORE	Open Research Europe
PIDs	Persistent IDentifiers
RO	Research Output
SAP	Self-Assessment Plan
SDM	Smart Data Model
SO	Scientific Objective
WDME	Water Data Management Ecosystem
WP	Work Package

1.0 INTRODUCTION

The first version of the Self-Assessment Plan (SAP) and the Data Management Plan (DMP) for the WATERVERSE project are described in this deliverable. The SAP introduces the appropriate strategies and indicators per scientific objective so as the performance of the WATERVERSE activities be monitored, and evaluated successfully, as well as be easily reported through the corresponding deliverables. This process will be supervised by WATERVERSE' Quality Assurance & Risk Manager (QARM) in closely collaboration with the Project Management Board (PMB), which is already described in the deliverable D1.1 and foreseen in the Grant Agreement.

Furthermore, the DMP is considered the formal document that outlines from the start of the project, every aspect of the research data lifecycle, which includes its organisation and curation, and adequate provisions for its access, preservation, sharing, and eventual deletion, both during and after a project[1] (Horizon Europe – Program Guide v2.0 11/04/2022, p43). Therefore, in this deliverable, a first version of DMP is included identified the datasets that will be handled during the WATERVERSE project. Specifically, it includes the procedures of acquisition, collection, processing and/or generating throughout the project's lifespan. It should be mentioned that this is a live document, in the sense that it will be periodically updated and extended as new datasets would be expected to employ via the project's activities. Hence, the second version of this deliverable, *D1.3 Self-assessment and data management plan v2*, is foreseen to publish at the 20th month (M20) of project's lifespan.

1.1 Purpose of the document

The D1.2 has dual purpose. Firstly, it aims to establish the Self-Assessment Plan (SAP) which indicates the processes to assess the WATERVERSE activities and tasks in terms of measurable and quantitative Key Performance Indicators (KPIs).

Secondly, it aims to specify the DMP which will support the data management life cycle for all data that will be gathered, processed, or generated by the WATERVERSE project. It will contribute to the efficient management of data in the project through the following steps:

- Outline the types of data collected and generated (or foreseen for collection and generation) during the activities of the WATERVERSE project.
- Describe the methodology and standards required, but also identify whether and how data will be collected, shared, exploited, re-used or made accessible for verification, and how they will be curated and preserved.
- Specify the degree of privacy and confidentiality of the collected/generated data.
- Outline the considerations and measures that are foreseen for the adequate management of the data from the legal, ethical, and security points of view.

[1] Horizon Europe – Program Guide v3.0, 01/04/2023 (pg. 43), https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf

- Outline the main elements of the data management policy that will be followed by the WATERVERSE consortium to handle collected/generated data with respect to their sensitiveness during and after the project.
- Ensure project research data and records are accurate, complete, authentic, interoperable and reliable.
- Enhance data security and thereby minimize the risk of data loss.
- Ensure research integrity and reproducibility by others.

The described policy reflects the current state of consortium agreements regarding data management and is consistent with those referring to exploitation and protection of results.

1.2 Structure of the document

The deliverable is structured as follows:

- In Section 2.0, a detailed Self-Assessment Plan is described upon the project scientific objectives as they have been declared in the Description of Actions (DoA).
- In Section 3.0, the data management structure is provided. It will be aligned with the during the WATERVERSE project according to the Horizon Europe – Program Guide v3.0 and the Guidelines on FAIR Data Management in Horizon 20202.
- In Section 4.0, the information about WATERVERSE datasets is provided. Initially, 33 datasets have been identified and described.
- In Section 5.0 conclusions and the future outlooks of the SAP and DMP are presented.

[2] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

2.0 ASSESSING PLAN

WATERVERSE aims to leverage innovative technology solutions that enhance data management practices and resources by developing a unified ecosystem, called Water Data Management Ecosystem (WDME). This holistic approach allows the accessibility, affordability, fairness, interoperability, usability, and security of the data and processes in the water domain. Also, it brings together competencies from the water domain (including water utility organisations, social sciences experts, etc.) along with technology providers and innovation companies from the technology community.

To achieve its mission WATERVERSE resolves specific Scientific Objectives (SO) and delivers concrete Research Outputs (RO), as described in the following sections (Section 2.1.1 and Section 2.1.2). Each SO consists of specific activities and assesses via the Key Performance Indicators (KPIs), that attempt to quantify and measure the performance of the particular SO. The Section 2.2 contains tables for the assessment of each Scientific Objectives including the evaluation strategy that will be used to assess the progress and quality of the performance of the SO as well as its KPIs. At the bottom section of each table, the expectations-target values for all the envisioned KPIs are provided.

2.1 Project Objectives

2.1.1 Scientific Objectives

SO1. Actively engage end-users and stakeholders to assess the main gaps and challenges the water sector must overcome to effectively be part of and contribute to quality European data spaces

- Particular attention will be paid to understanding behaviours, motivations, and barriers, with the goal of maximizing the uptake of WATERVERSE outputs. We will organise **Multi-Stakeholder Forums (MSF)** and provide evidence in a comprehensive way to **stimulate behaviour change** of **stakeholders** (**utilities**, **policy makers**, **citizens**) towards data management and data sharing/exchange for the water sector and better use of the data spaces for water across Europe.
- Critically assess today's **real-world data spaces in the water sector**, highlighting the **gaps and needs from various perspectives**, including political, social, economic, and technical, and identify or define new opportunities and scenarios to overcome current barriers through the incremental development of data spaces based on FIWARE models, on Open Standards/Specifications and on results of Gaia-X.
- Analyse and characterize **real data management and heterogeneity challenges** in diverse use cases of the water sector, specifying technical requirements from the phases of the data management lifecycle.
- Reduce **data barriers** caused by the different data governance frameworks at EU and international level that impede **data sharing**, **sovereignty** and **accessibility**.

SO2. Identify, extend, and integrate a wide set of data management tools to implement the WDME, based on FIWARE (www.fiware.org) Building Blocks

- Create an extendable, distributed, and user-friendly ecosystem for data management, composed of tools and data resources that can be used independently or in combination through a **practical GUI-based pipeline approach** that allows the tool ecosystem to **scale from simple tasks to complex operations** that need multiple tools to be pipelined.
- Build upon data space **components and assets already available from FIWARE and IDSA** to ensure **full compatibility with European data spaces** and provide the complementary tools and resources to enable effective data management practices for efficiently feeding data spaces in the water sector.

- Populate the data management tool ecosystem with an initial **set of robust tools that result from previous projects of the consortium members and beyond**, spanning from data discovery and collection to data preparation and meta-characterisation, including:
  1. Tools and procedures for **supporting the creation and maintenance of common data models, in the framework of the Smart Data Model initiative**.
  2. **AI-based semantic annotation** of domain data to build the Water Sector Data Graphs.
  3. **Water ontologies** and semantic models
  4. **Data fairness** by a set of AI-based tools that permit to **create balanced and fair datasets**.
  5. **Data quality** and **integrity** improvement through AI-based tools for **data validation**, **anomaly detection** and **data reconciliation/correction**.
  6. Data **anonymization**.
  7. **Blockchain-based data provenance and inviolability** of data management.
  8. **Synthetic IoT data generation**.

SO3. Setup and demonstrate the WATERVERSE WDME in real environment with relevant and diverse case studies involving water sector stakeholders from 6 countries (the Netherlands, Germany, Cyprus, United Kingdom, Spain, Finland)

- Implementation, monitoring, and evaluation plans for **co-designed data sharing** processes in each demonstration area, including the definition of key indicators to identify, analyse, and quantify (where relevant) the **economic/social/environmental benefits and trade-offs** of the water sector data spaces.
- Implementation and demonstration of the **WATERVERSE WDME** through **two-iterations of specific actions in 6 diverse pilot areas**, with their actual data management operations and their engaged end-users and stakeholders, namely in Netherlands, Germany, Cyprus, United Kingdom, Spain, and Finland.
- Provide data management guidelines and recommendations tailored to the water sector, with focus on **business opportunities for SMEs and start-ups**, and how they can leverage the results of the project for lowering the entry barrier to European and international data spaces and to the overall data economy.

SO4. Ensure security and energy efficiency of the WATERVERSE WDME

- Integrate advanced cybersecurity solutions to guarantee **Confidentiality**, **Integrity**, and **Availability** (**CIA triad**) on the data management systems, with a focus on **threat detection**, **integrity monitoring**, and **incident response**, also adopting **Cyber Threat Intelligence** (**CTI**) techniques to enhance the capabilities of Intrusion Detection Systems.
- Establish a precursor, **energy-efficient approach** based on machine-learning technologies to serve as a basis for further machine-learning approaches (e.g., classification/categorization, anomaly/outlier detection, root-cause analysis, and predictive analytics) which can be applied in the centralized/distributed data management solutions.

SO5. Set clear and measurable indicators for assessing FAIRness of data in water-related data spaces

- Define the concepts for **FAIR Digital Objects** and **FAIR Ecosystems** which include data as well as related services and infrastructures.
- Deliver **guidelines**, **recommendation**, **metrics**, and **tools** to generate assessment reports of the **FAIR maturity level** related to FAIR Digital Objects and FAIR Ecosystems in the **water sector**.
- Provide **contextualized measures** of **FAIRness** at the beginning and at the end of pilot cases for the relevant data, data management tools, services, and infrastructures.

- **Go beyond** FAIR principles and apply the principles of **MELODA5** to calculate the dissemination and reputation of the water service, data, and infrastructures.
- Analysis of other MELODA5 principles to be adopted in the project to **improve the open data reusability**.
- **Elaborate AI-driven tools** that permit to balance datasets also avoiding data discrimination. At the end, we will create **trustworthy** and **transparent** data strategy to ensure fair and non-discriminatory datasets.

SO6. Ensure the viability and sustainability of the WATERVERSE WDME, as well as its replicability, scalability and business applicability

- Deliver a workable **exploitation and business plan**, to promote the **adaptation**, **replicability**, **scalability**, and **exploitation** of the WATERVERSE WDME and for water utilities, technology companies, research organisations, and other end-users across Europe and beyond. Deal with market analysis, IP and innovation management, business risks, financial models, and projection.
- Provide **policy and governance recommendations** to support the widespread adoption of the WATERVERSE WDME and approach, as well as the business applicability of the water data spaces.
- Perform dedicated **communication, dissemination, outreach**, and **clustering activities** to promote and increase the impact of the WATERVERSE WDME.

2.1.2 Research Outputs

The above SOs will be substantiated through a series of ROs, which can be considered as WATERVERSE's main technological solutions that will be developed and integrated into the WDME. A short description of the objectives of each RO is provided below:

RO#1 – Data catalogue for Open Data discovery: its objective is to federate open data so that there is a single point of access and keep track of it in a catalogue. This enables users to have a single point of access to Open Data, keep it up-to-date, and discover new ones through Web Scraping.

RO#2 – IoT Agents and system adapters: its objective is to integrate any Context Information Provider into Powered-by-FIWARE Architectures. An IoT Agent translates an IoT proprietary protocol and transport mechanism into NSGI (v2 or LD) to communicate with the FIWARE Context Brokers.

RO#3 - Smart Data Models for data harmonization: its objective is to Harmonize the data provided by the Use Cases to enable data portability to different applications. Smart Data Models (SDMs) allow actual data interchange between organizations by providing open licensed shared data models according to the principles of agile standardization.

RO#4 - IoT Device Manager: its objective is to register "push" and "pull" IoT devices, gather data from them and store it in the WDME. With the help of RO#2, it will be possible to register sensors and collect data using different IoT protocols.

RO#5 - Data Preparation Pipeline Editor: its objective is to enable the processing, harmonization and integration of data in a no-code mode, through a purely graphical approach. Specifically, the main goal of the tool is to integrate all the tools that collect, generate, and process data within a single environment, so that there is a single point of access and even non-expert users can graphically draw data streams.

RO#6 - Artificial Intelligence (AI) Based Data Validation Tool: its objective is to automate the data quality control processes, to identify anomalies (such as missing values, outliers, irregular trend breaks, drifts etc.) in crucial datasets and to provide standardized and interoperable solutions through integration with FIWARE.

RO#7 - Synthetic IoT data generator: its objective is to provide large-scale (potentially) plausible 'normal' data, to supply edge-case data on-demand removing the need to develop tedious code from scratch.

RO#8 – WaterAnalytics digital twin as an EPANET distribution supported by Aqualogist Web Platform: its objective is to provide a software platform offering IoT integration (sensors, actuators, other processing functions) and facilitating the real-time monitoring and control of large-scale systems. The platform utilises PHOEBE's SEMIoTICS architecture, which supports the data semantic annotation and reasoning models. The WaterAnalytics solution uses the EPANET tool as its front-end, with a dedicated distribution that integrates telemetry, water network models and smart algorithms that facilitate the decision-making process of water operators.

RO#9 - Data anonymization tool: its objective is to convert personal data to non-personal data, irreversibly, with minimized uniqueness of combinations of records, and null re-identification risks.

RO#10 - Data clusterability tool: its objective is to measure the clusterable structure in the given data set to pre-evaluate the success of clustering analysis. Also, it inspects the quality of the data (structured/noise) for applicability of further analyses (e.g., unsupervised classification, refined outlier detection, and predictive analytics).

RO#11 – Blockchain–based data provenance tool: its objective is to support the persistence and verification of NGSI-LD Entity-Transactions in blockchains. Specifically, it permits the registration of any transaction over a dataset as well as any operation applied over it together with the identification of the user/Organization and requests the list of transactions developed in the original dataset.

RO#12 – Tools for FAIRness assessment: its objective is to provide a set of tools able to facilitate the creation of FAIR+MELODA5 structured data models. The application of those tools allows us to evaluate and create data models associated with the Use Cases taking into account FAIR Principles and MELODA5 dimensions.

RO#13 - Energy efficiency of the data management tools: its objective is to improve energy efficiency throughout the data management pipeline. The energy saving begins from the physical layer with optimizing sensor readings. In the middleware layer, the tool is expected to leverage edge computing together with data preprocessing, aggregation, and transfer optimization. In the application layer, the tool aims to minimize storage requirements and computation times of the user applications used to analyze and visualize the data.

RO#14 - Cybersecurity solutions for Water Utilities: its objective is to provide a set of tools that guarantee the Confidentiality, Integrity and Availability of Water Utility systems by leveraging Cyber Threat Intelligence (CTI) and utilising it in the domain of *Intrusion Detection* and *Incident Response*.

RO#15- Data exploration and navigation tool: its objective is to develop an interactive semantic knowledge graph. This tool will integrate information from different semantic data-stores and graph-data stores, enabling data exploration and navigation from different perspectives through semantic relationships.

RO#16- Water Ontology Catalogue: its objective is to develop a set of semantic models based on the SAREF4WATR standard model to facilitate data harmonization and metadata management in the domain of water management. This involves creating ontologies for different domains, including waste-water, risk assessment, water quality, and water-energy-resource symbiosis domains.

RO#17- Data Balancing and Non-discriminatory Algorithms: its objective is to develop a framework that comprises the collection of data quality algorithms to automatically remove outliers, reduce data noise, remove missing values (data blanks). Also, it encapsulate algorithms for class balancing and non-discriminator.

RO#18- FIWARE Data Converters: its objective is to provide a set of tools able to transform unstructured data into structured NGSI – LD data. Hence, the data will be available in a structured way, harmonizing the data acquisition process and making access to those data easier and more convenient.

For the efficient monitoring of the assessment process, a mapping of each SO with specific WPs and related tasks is carried out and presented in the following table. This association will permit us to evaluate the progress and performance of the provided solutions and ROs via the achievements of the corresponding tasks.

SO#	SO Description	ROs	Relative WPs/Tasks
SO1	Actively engage end-users and stakeholders to assess the main gaps and challenges the water sector must overcome to effectively be part of and contribute to quality European data spaces.		WP2 (T2.1, T2.2, T2,3) WP5 (T5.1)
SO2	Identify, extend, and integrate a wide set of data management tools to implement the WDME, based on FIWARE (www.fiware.org) Building Blocks.	RO#1 RO#2 RO#3 RO#4 RO#5 RO#6 RO#7 RO#8 RO#9 RO#10 RO#11 RO#15 RO#16 RO#17 RO#18	WP2 (T2.4) WP3 (T3.1, T3.2, T3.5)
SO3	Setup and demonstrate the WATERVERSE WDME in real environment with relevant and diverse case studies involving water sector stakeholders from 6 countries (Cyprus, Spain, Germany, the Netherlands, Finland, United Kingdom)		WP2 (T2.1, T2.2) WP5 (T5.1 - T5.3)
SO4	Ensure security and energy efficiency of the WATERVERSE WDME.	RO#13	WP3 (T3.3, T3.4)
SO5	Set clear and measurable indicators for assessing FAIRness of data in water-related data spaces	RO#3 RO#12	WP4 (T4.1 - T4.3)
SO6	Ensure the viability and sustainability of the WATERVERSE WDME, as well as its replicability, scalability and business applicability		WP2 (T2.5) WP6 (T6.1 - T6.5)

Table 1: Mapping between SOs, ROs and relative WPs/Tasks

2.2 Self-assessment plan per Scientific Objective

2.2.1 Scientific Objective 1

<table>
<tr><th colspan="2">Scientific Objectives</th><th>KPIs</th></tr>
<tr><td colspan="2" rowspan="3"><i>SO1. Actively engage end-users and stakeholders to assess the main gaps and challenges the water sector must overcome to effectively be part of and contribute to quality European data spaces</i></td><td>KPI 1.1: Number of data sharing processes affected by the gaps/challenges</td></tr>
<tr><td>KPI 1.2: Number of gaps/challenges identified</td></tr>
<tr><td>KPI 1.3: Number of involved types of stakeholders per Case Study in MSF</td></tr>
<tr><th colspan="3">Evaluation Strategy Description</th></tr>
<tr><td colspan="3"><ul><li>Continuously monitoring the processes for the developing of the adequate framework for stakeholders’ engagement (e.g. end-users, policy makers) in data spaces and data management.</li><li>Monitoring and evaluating the processes for identifying stakeholders, setting up Multi-Stakeholder Forums (MSF) with topics and planning roadmap, and facilitating the</li><li>dialogue process.</li><li>Monitoring and accessing the progress of the 3 phases of the MSF meetings at the case studies.</li><li>Every WP2 task/activity will be continuously monitoring in order to ensure that the outcomes will be in accordance with the given timelines especially on the technical perspective.</li></ul></td></tr>
<tr><th colspan="3">Indicators</th></tr>
<tr><td><b><i>#</i></b></td><td><b><i>Highest expectations</i></b></td><td><b><i>Lowest expectations</i></b></td></tr>
<tr><td>KPI 1.1</td><td>12+ data sharing processes</td><td>At least 10 data sharing processes</td></tr>
<tr><td>KPI 1.2</td><td>12+ identified gaps/challenges</td><td>At least 10 identified gaps/challenges</td></tr>
<tr><td>KPI 1.3</td><td>12+ stakeholders per case study</td><td>At least 8 stakeholders per case study</td></tr>
</table>

Table 2: Self-Assessment Plan of SO1

2.2.2 Scientific Objective 2

<table>
<tr><th>Scientific Objectives</th><th>KPIs</th></tr>
<tr><td rowspan="3"><i>SO2. Identify, extend, and integrate a wide set of data management tools to implement the WDME, based on FIWARE (www.fiware.org) Building Blocks</i></td><td>KPI 2.1: Number of standard data models supported for the water sector</td></tr>
<tr><td>KPI 2.2: Data collection/discovery tools/resources integrated</td></tr>
<tr><td>KPI 2.3: Data preparation tools integrated</td></tr>
<tr><th colspan="2">Evaluation Strategy Description</th></tr>
<tr><td colspan="2"><ul><li>Identify the open data sources and existing Smart Data Models dedicated to the water sector.</li></ul></td></tr>
</table>

• Monitoring the development of the tools/resources for data discovery, data collection and data preparation by considering the functional and non-functional requirements specifying in WP2 and reported in D2.1. • Continuously assess the performance of data discovery/collection/preparation tools using specific quantitative indicators and metrics. • Evaluating the integration processes of the data discovery/collection/preparation tools/resources to the WDME.		
Indicators		
#	***Highest expectations***	***Lowest expectations***
KPI 2.1	All Smart Data Models under Water Domain of the Smart Data Initiative	At least the Smart Data Models that are needed by pilot will be supported for the water sector
KPI 2.2	All the tools/resources for data discovery/collection generated by RO#1, RO#2, RO#4, RO#16, RO#9, and additional tool needed to collect different types of data will be integrated	At least RO#1, RO#2, RO#4 will be integrated
KPI 2.3	Tools for data preparation generated by RO#5 to RO#11, RO#15, RO#17, RO#18, and additional tools that emerge during the project will be integrated	At least tools for data preparation generated by RO#5 to RO#11, RO#15, RO#17, RO#18 will be integrated
Research Outputs		
KPI 2.1	RO#3 - Smart Data Models for data harmonization RO#16 - Water Ontology Catalogue	
KPI 2.2	RO#1 - Data catalogue for Open Data discovery RO#2 – IoT Agents and system adapters RO#4 - IoT Device Manager	
KPI 2.3	RO#5 - Data Preparation Pipeline Editor RO#6 - Artificial Intelligence (AI) Based Data Validation Tool RO#7 - Synthetic IoT data generator RO#8 – WaterAnalytics digital twin as an EPANET distribution supported by Aqualogist Web Platform RO#9 - Data anonymization tool RO#10 - Data clusterability tool RO#11 – Blockchain–based data provenance tool RO#15 - Data exploration and navigation tool RO#17 - Data Balancing and Non-discriminatory Algorithms RO#18 - FIWARE Data Converters	

Table 3: Self-Assessment Plan of SO2

2.2.3 Scientific Objective 3

Scientific Objectives	KPIs
SO3. Setup and demonstrate the WATERVERSE WDME in real environment with relevant and diverse case studies involving water sector stakeholders from 6 countries	KPI 3.1: Number of end-user organisations directly engaged in demonstrations
	KPI 3.2: Number of use-cases addressed and enabled by the WATERVERSE WDME
	KPI 3.3: Performance indicators (detailed in Task 5.3) reaching or exceeding targets. ***Economic and Environmental Impact Assessment metrics:*** • Reduce the time required to execute a service/procedure. • Increase the learnability and usability of a service/procedure ecxecution. • Reduce the cost of executing a service/procedure. • Improve the quality of the offered services/procedures. • Increase the overall efficiency of the organisations' operations. • Increase the efficiency of the resources' usage. • Minimise the energy consumption for executing a service/procedure. • Optimise the environmental footprint of a service/procedure. ***KPIs:*** • Number of procedures/services enabled • Time required to run a procedure/service that involves data management • Quantified improvement per each service, e.g., 10% reduction of drinking water leakages, 15% reduction in consumer complains. • Quality improvement of a service, e.g., higher satisfaction of the water consumers, water quality within appropriate bounds, etc. • Cost of execution of a certain service/procedure. • Energy consumption of a service/procedure • Overall greenhouse gas emissions of a service/procedure. • Personnel involvement in executing a service/procedure.

<table>
<tr><td colspan="2"></td><td>KPI 3.4: Number of external end-users and stakeholders involved in each pilot</td></tr>
<tr><th colspan="3">Evaluation Strategy Description</th></tr>
<tr><td colspan="3">• Prepare the pilot sites in collaboration with each of the pilot partners.
• Define a detailed pilot/demonstration plan to accommodate for two pilot iterations to evaluate the WATERVERSE WDME.
• Define operation scenarios and use-cases for each site, taking into consideration the needs, end-users' requirements and expertise of the system operators.
• Deploy the selected individual WATERVERSE WDME tools in each of the pilot sites.
• Train the pilot partner's operators at each pilot site who will be working with the WATERVERSE Platform during pilots.
• Evaluate the operational usability and technical completeness of the solution and the degree to which it comprises a "pain killer" for the everyday pain of the organisations in relation with their needs to use data and the cost of doing so.
• Collect feedback (structured questionnaires, as well as semi-structured interviews) through an on-line platform.
• Analyse the impact in improving the operations of the pilot organisations.</td></tr>
<tr><th colspan="3">Indicators</th></tr>
<tr><th><i>#</i></th><th><i>Highest expectations</i></th><th><i>Lowest expectations</i></th></tr>
<tr><td>KPI 3.1</td><td>6 end-users organisations will be directly engaged in demonstrations</td><td>At least 1 organisation per Pilot Use Case will be involved in the execution of the 1st and 2nd iteration of the pilots</td></tr>
<tr><td>KPI 3.2</td><td>More than 3 UCs per pilot (i.e., more than 18 UCs in total).</td><td>3 UCs per pilot, i.e., 18 UCs in total</td></tr>
<tr><td>KPI 3.3</td><td>Any achievement higher than the minimum targets set per pilot site, will be considered, and recorded. There are no strict highest expectations related to this set of KPIs.</td><td>In the framework of the pilot planning, each pilot site will define minimum targets for the pre-defined KPIs/metrics. These minimum targets will depend on the specifics of the pilot UCs and the involved technological tools.</td></tr>
<tr><td>KPI 3.4</td><td>More than 48 organisations involved in general.</td><td>At least 48 external end-users and stakeholders involved in the WATERVERSE WDMS pilots (at least 8 per pilot site).
It is not expected for all these organisations to participate actively in the demonstrations. However, it is expected that they will get engaged in various ways.</td></tr>
</table>

Table 4: Self-Assessment Plan of SO3

2.2.4 Scientific Objective 4

Scientific Objectives	KPIs
SO4. Ensure security and energy efficiency of the WATERVERSE WDME	KPI 4.1: Sub-quadratic (< O(n^2)) processing time for the precursor approach
	KPI 4.2: Mean time to detect attacks within the system
	KPI 4.3: Mean time WDME is not fully operational for cybersecurity incident
	KPI 4.4: Mean time to acknowledge vulnerability or threat (based on the collected CTI)
	KPI 4.5: Mean time to patch a vulnerable system
	KPI 4.6: Reduction of computation time and data volume

Evaluation Strategy Description
• Create questionnaire for the Use Cases about the cybersecurity solutions and practises they use. • Collect logs from end-users and from honeypots. • Re-evaluate AI models and tailor them according to the needs of water utilities. • Research and collect relevant external sources (e.g. CERT feeds, relevant vulnerability databases). • Explore, implement, and evaluate open-source IDS/IPS and Incident Response solutions. • Establish and Evaluate energy-efficient algorithms and solutions aiming to minimize energy footprint of the most demanding WATERVERSE tools in terms of storage, computation and network usage.

Indicators		
#	***Highest expectations***	***Lowest expectations***
KPI 4.1	Sub-quadratic (< O(n^2)) processing time for the precursor approach	Quadratic (O(n^2)) processing time
KPI 4.2	10% decrease over the baseline	5% decrease over the baseline
KPI 4.3	10% decrease over the baseline	5% decrease over the baseline
KPI 4.4	15% decrease over the baseline	5% decrease over the baseline
KPI 4.5	15% decrease over the baseline	7% decrease over the baseline
KPI 4.6	15% reduction in energy demand through faster computation and numerosity reduction	10% reduction in energy demand through faster computation and numerosity reduction

Research Outputs	
KPI 4.1 – KPI 4.5	RO#14 - Cybersecurity solutions for Water Utilities
KPI 4.6	RO#13 - Energy efficiency of the data management tools

Table 5: Self-Assessment Plan of SO4

2.2.5 Scientific Objective 5

Scientific Objectives		KPIs
SO5. Set clear and measurable indicators for assessing FAIRness of data in water-related data spaces		KPI 5.1: Number of Smart Data Models in the Water Domain compliant with FAIR principles
		KPI 5.2: FAIR Digital Objects (WDME services) fully described and adopted in the different Use Cases of the project
		KPI 5.3: Number of MELODA5 dimensions adopted in our guidelines
Evaluation Strategy Description		
• Define the WATERVERSE FAIR Guidelines for the water sector taking into account the FAIR and MELODA5 principles, based on the different level of data sharing, data types as well as the proper nature of the research of each stakeholder. • Identify the corresponding keys where the FAIR and MELODA5 principles will need to be expanded and unpacked during the execution of the project. • Define what should be needed to be FAIR a FAIR Digital Objects (services) and the corresponding components put in place to deploy a FAIR Ecosystem. • Develop FAIR Metrics for FAIR Digital Objects and FAIR Ecosystem. • Define and evaluate the Smart Data Models in the Water Domain compliant with FAIR principles. • Implement the FAIR Data Management Plan.		
Indicators		
#	***Highest expectations***	***Lowest expectations***
KPI 5.1	More than 12 Smart Data Models in the Water Domain that will be complaint with FAIR principles	At least 6 Smart Data Models (one per each pilot) in the Water Domain that will be complaint with FAIR principles
KPI 5.2	More than 8 FAIR Digital Objects (WDME services) fully described and adopted in the different Use Cases of the project	4 FAIR Digital Objects (WDME services) fully described and adopted in the different Use Cases of the project
KPI 5.3	All MELODA5 dimensions adopted in our guidelines	2 MELODA5 dimensions adopted in our guidelines
Research Outputs		
KPI 5.1 – KPI 5.3	RO#3 - Smart Data Models for data harmonization RO#12 – Tool for FAIRness assessment	

Table 6: Self-Assessment Plan of SO5

2.2.6 Scientific Objective 6

Scientific Objectives	KPIs
SO6. Ensure the viability and sustainability of the WATERVERSE WDME, as well as its replicability, scalability and business applicability	KPI 6.1a: Number of delivered governance/policy recommendations involving gaps/barriers identified in WP2
	KPI 6.1b: Number of delivered recommendations for the water data spaces
	KPI 6.2a: Number of potential purchasers (water utility organisations)
	KPI 6.2b: Number of potential WATERVERSE clients who will be participated in the business meetings organised by the project

Evaluation Strategy Description
• Define and adopt a concrete policy and governance recommendations for measures (support initiatives/ funding schemes/ subsidy schemes) at the country level. • Assess and analyse WATERVERSE outcomes to support the decisions of the consortium regarding the commercial exploitation channels. • Define a clear set of activities, targeting the active engagement of stakeholders, potential customers and end-users. • Setting the foundations for protecting and sustaining the product beyond the duration of the project. • Address all market and business aspects through the WDME. • Continuously monitoring, through internal 6-month reporting from the consortium's partners, the number of networking activities achieved, the audiences reached, and the impact created. • Continuously monitoring, through internal 6-month reporting from the consortium's partners, the progress of the communication and dissemination actions. • Continuously assess the successful organization and implementation of the planned conferences, workshops and info-days.

Indicators		
#	***Highest expectations***	***Lowest expectations***
KPI 6.1a	More than 6 governance/policy recommendations (1 per Case Study) involving gaps/barriers identified in WP2 and recommendations for the water data spaces	At least 6 governance/policy recommendations (1 per Case Study) involving gaps/barriers identified in WP2
KPI 6.1b	More than 6 recommendations (1 per Case Study) for the water data spaces	At least 6 recommendations (1 per Case Study) for the water data spaces
KPI 6.2a	Receive intention to explore purchases by 6 organisations (water utilities)	Receive intention to explore purchases by at least 3 organisations (water utilities)

KPI 6.2b	More than 10 potential WATERVERSE clients, to whom the business model will be presented and explained, and will be participated in project's business meetings	At least 10 potential WATERVERSE clients, to whom the business model will be presented and explained, and will be participated in project's business meetings
KPI 6.3	Attract more than 1000 members of key stakeholders in events and exploitation activities (coorganised) by WATERVERSE	Attract at least 1000 members of key stakeholders in events and exploitation activities (coorganised) by WATERVERSE

Table 7: Self-Assessment Plan of SO6

3.0 DATA MANAGEMENT STRUCTURE

3.1 Data Summary

Under this section an overview of the datasets identified by WATERVERSE partners as being collected and generated during the project at the current stage. This summary and the datasets per se will be refined, supplemented, and expanded, during the lifespan of the project, making the Data management Plan as a living document. Updated versions of the DMP will be included in subsequent internal reports. All the datasets have been collected by the partners using the Horizon Europe Data Management Plan Template[3] as a basis, being adapted to the needs of the WATERVERSE project and the requirements as they have been formed till M7 (check Section 3.2 and ANNEX A for the WATERVERSE template). Partners have also reported whether personal data would be included in their datasets, with additional information to have been outlined in the Personal Data protection (Section 3.5). The following sections illustrate the sources and type of data, a general description of datasets (with the specifications of datasets to be included in Section 4.0), along with information on the data management in the technical and non-technical related work of WATERVERSE project.

The main principles described in this document are expected to remain relevant until the end of the project, with WATERVERSE aiming to facilitate the use and reuse of project data while ensuring compliance with legal and ethical considerations through the development and implementation of effective data management strategies and practices.

3.2 WATERVERSE Data: Purpose of Data collection/generation and Reporting methodology

The purpose of the data collection/generation and processing in WATERVERSE project is mainly to fulfil the aims and objectives of the project in relation to the use-cases implementation and piloting, as well as to the project management and to the communication dissemination and exploitation activities. More in particular:

- **Project Management data**: Having as main objective the proper management and coordination of the project, the data that will be collected are:
  a. Contact data from all partners collected for internal Consortium communication (WP1)
  b. Contact data from the relevant Advisory Board (T1.1)
  c. Deliverables, Meeting Minutes, Handbooks Official Reports and Review Documents, Financial data (budget etc.) (All WPs)
  d. Project management meetings through digital communication (email, video conference, etc.) (WP1)
  e. Contact data (names, contact details, professional affiliation) as well as photos, videos and audio recordings from relevant stakeholders that will form the WATERVERSE stakeholders' group (T2.1)

[3] https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx

- **Communication and Dissemination data**:
  f. Contact data/lists from networking synergies, clustering, and outreach activities (T6.3)
  g. Events and Scientific publications datasets (WP6)
  h. Website and newsletter subscribers (WP6)

- **Technical Development, Research and Piloting Data:**
  i. Data stemming from the stakeholders' survey for end user requirements (online questionnaire, transcripts from workshops, etc.) (WP2)
  j. Data stemming from policy recommendations documents (WP2)
  k. Information extracted from the pilot implementation, as regards feedback and evaluation questionnaires (online questionnaire, transcripts from workshops, etc.) (WP5)
  l. Primary and Secondary datasets (from SCADA, sensors, open data energy/water platforms, etc.) used, among other, to generate the relevant KPIs regarding the efficiency of the network and the customers behaviour, to facilitate the modelling, visualisation, to perform certain data exchange tests, to populate the threat intelligence module, to validate the raw sensor data and screening for anomalies as well as to train the relevant ML-models (WP3, WP4, WP5). These datasets will be further enhanced to facilitate additional tools of WATERVERSE project.

To ensure the protection of personal data, additional information about how this information will be managed is included in Section 3.5.

A specific DMP template, is included in the ANNEX A, was distributed among partners, in an excel format (here converted in word for formatting issues). All consortium partners have filled in the relevant information listing and detailing the datasets they have identified that will be suitable for the projects' objectives. The master document, (WATERVERSE DMP Master document), is a living document stored in the Project Repository and will be used as a reference for all datasets used in WATERVERSE throughout the project's lifespan. Partners are responsible for updating the document whenever a new dataset emerges or there are updates on a listed one. The full list of datasets (Datasets information) provided by partners is in Section 4.0.

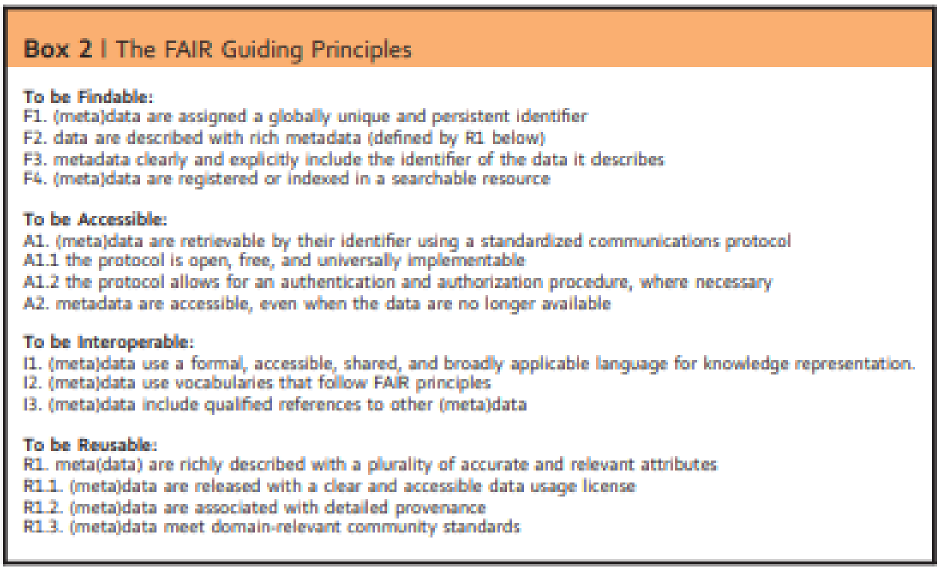
3.3 FAIR data

This section outlines on a high-level, all the measures that will be taken to ensure that the data collected will be Findable, Accessible, Interoperable, and Reusable (FAIR). More information and a detailed analysis of the FAIR data principles in WATERVERSE project will be further identified, analyses and reported in detail under *WP4 - FAIRness assessment in the water industry* and its relevant deliverables D4.2, D4.4 and D4.6.

The acronym FAIR (Findable, Accessible, Interoperable, and Reusable) has been initially defined in 2016[4], with the following Figure stemming from the said paper to accurately summarize the relevant principles. As it has been underlined in the said paper, FAIR principles should apply to all types of data, including "*algorithms, tools, and workflows that led to that data*", enabling in that way knowledge discovery and innovation, and support scientific reproducibility and replicability. FAIR principles have been widely communicated by the European Commission (EC) in the Communication "European Data Strategy (2020)", while Horizon Europe puts a great emphasis on the "*management of data and other outputs in*

[4] Wilkinson MD, et.al, 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 15;3:160018. doi: 10.1038/sdata.2016.18.

accordance with the FAIR principles, which means making data Findable, Accessible, Interoperable, and Reusable”[5].



Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Figure 1: FAIR Acronym and Guiding principles (from Wilkinson MD, et.al, 2016[4])

Since the WATERVERSE project is publicly funded, it is important to follow the FAIR principles for research data, wherever applicable. This plan aims to extend these principles to the datasets generated and collected during the innovation activities of the project. The following sub-sections provide a high-level description on the way WATERVERSE Project will implement the FAIR principles, while a more detailed and robust analysis will be performed in WP4 and in its relevant deliverables.

3.3.1 Making Data Findable

WATERVERSE project aims to comply with the Horizon Europe Open Access requirements[6], using the Zenodo repository where all publicly accessible datasets, scientific publications, and deliverables will be uploaded, thus ensuring in that way that research data is findable. The research data will be linked to the OpenAIRE community and/ or Open Research Europe (ORE) platform community[7] to ensure its maximum visibility and availability. The usage of Persistent Identifiers (PIDs), such as Digital Object Identifiers (DOIs)[8] will be sought to facilitate data citation. Clear and proper file naming, with the minimum possible number of characters that sufficiently describes the content will be also adopted. This DMP proposes an initial list of good practices to partners to follow, which will be further enhanced and finalised in WP4 and its relevant deliverables:

- No special characters: @, #, etc.
- No white space, instead underscore is recommended
- No abbreviations or codes
- Use of period only before the file –name extensions

[5] https://www.openaire.eu/how-to-comply-with-horizon-europe-mandate-for-rdm
[6] Horizon Europe, open science - Publications Office of the EU (europa.eu)
[7] Open Research Europe | Open Access ... | Open Research Europe (europa.eu)
[8] https://www.doi.org/

- File version to be included
- Standard date following ISO (8601) (YYYYMM)

File naming includes rules standard:

- Standard date following ISO (8601) (YYYYMM)
- Project name
- File Name
- Main-Author
- Version number

In addition, each WATERVERSE partner will have the responsibility of assigning specific keywords to all publicly uploaded datasets on Zenodo. These keywords should accurately depict the contents and nature of the datasets.

As far as Metadata is concerned, FAIR principles will be also applicable to them, taking into consideration the Dublin Core Metadata Initiative[9] along with the DataCite Metadata Schema 4.4 (Released in March 2021)[10], taking into consideration or the relevant updates. Generally, the metadata of every uploaded element on Zenodo is immediately indexed, included, and made searchable on Zenodo's search engine upon publishing. WATERVERSE consortium has generally incorporated a high-level availability of metadata in the datasets they aim to collect, with a more comprehensive information about dataset metadata to be outlined in detail and added both to the relevant WP4 deliverables and to the next version of this plan.

3.3.2 Making Data Accessible

The primary objective of the Horizon Europe Open Access requirements is to facilitate access to research data produced through Horizon Europe projects, aiming to build on previous results, encourage collaboration, speed up innovation and keep citizen and society involved. In alignment with the EC Guidelines on Open Research Europe (ORE) platform – Open Science[11], and following the Model Grant Agreement for the Horizon Europe Program, WATERVERSE will adopt an open access dissemination policy for the published research work, through Zenodo, which is a general-purpose repository to share open and FAIR research outputs, and directly linked with OpenAIRE platform. It will also ensure open access to peer-reviewed scientific publications, depositing the data in a trusted repository and adopting “the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a licence with equivalent rights, following the principle ‘as open as possible as closed as necessary’”[12]. However, it should be noted that all data will be, prior to being made publicly available, properly scrutinized, according to privacy, legal, ethical, and regulatory requirements. Parallel to that, all legitimate interests of beneficiaries including commercial exploitation along any relevant constraints will be also taken into consideration, being properly detailed in the second version of DMP all the legitimate exception(s) under which they choose to restrict access to (some of the) research data (if emerged). Access to confidential data, such as non-anonymous datasets that raise

[9] https://www.dublincore.org/specifications/dublin-core/dces/

[10] https://schema.datacite.org/meta/kernel-4.4/

[11] Horizon Europe, open science - Publications Office of the EU (europa.eu) and programme-guide_horizon_en.pdf (europa.eu)

[12] https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/agr-contr/unit-mga_he_en.pdf

security or privacy concerns, as well as innovation data that is intended for commercial and/or exploitation purposes, will be restricted to WATERVERSE partners.

3.3.3 Making Data Interoperable

WATERVERSE Consortium aims to make the data and tools they will create interoperable through adopting common standardised file open formats and standards and controlled vocabularies. WATERVERSE, a mentioned above, will utilise Zenodo during its lifespan, using standard vocabularies and machine-readable file formats, thus ensuring in that way interoperability. Particularly, to represent internal metadata, Zenodo employs JSON schema and supports open formats such as Dublin Core, MARCXML, BibTeX, CSL, DataCite, and Mendeley. The vocabularies used in the data record metadata will be consistent with those used by Zenodo. For data use vocabularies, Zenodo refers to open external vocabularies, such as Open Definition for licenses, FundRef for funders, and OpenAIRE for grants. External metadata is referenced using a resolvable URL[13]. Initial information on interoperability has been collected and discussed among partners, with the final decision to be outlined in detail and added both to the relevant WP4 deliverables and to the next version of this plan.

3.3.4 Making Data Reusable

In order to make data reusable, proper documentation, clear licencing (e.g., Creative Commons Licenses (CCL) or other open-source initiative licenses such as MIT License or Apache2.0 License) and provenance information has to be ensured. WATERVERSE will assess the specific data (and metadata) licences on a case-by-case basis in close collaboration with the involved WATERVERSE project partners, taking into account the legal and ethical requirements, in order to select the appropriate licensing scheme. Data and metadata will be sought to be openly available, unless restricted by confidentiality or data protection. The research data generated by WATERVERSE will be anonymized and subject to the Creative Commons Licenses applied to the shared data on Zenodo for third-party reuse. All publicly available research data stored in the main Zenodo repository are intended to remain usable indefinitely[14]. In addition to data stored in the Zenodo repository, publishable (non-confidential) data will be accessible for verification and reuse in WATERVERSE project repositories, accessible through the project website, for up to five years beyond the project. As in all other FAIR principles, initial information on data re-use has been collected and discussed among partners, with the final decision to be outlined in detail and added both to the relevant WP4 deliverables and to the next version of this plan.

3.3.5 Allocation of Resources

Finally, as far as the estimated costs of making the data FAIR, each entity has the responsibility to manage its own data, but CERTH and the leader of the Data Management task will supervise data management, covering the practical implementation of the DMP. Partners are responsible for data generation, metadata production, and data quality, with specific responsibilities depending on the data and the internal organization in the work packages and tasks where data is created or used.

All estimated costs associated with the process, are integrated in the project’s budget, covering specific data processing and management activities, from collection and documentation to storage, preservation,

[13] https://about.zenodo.org/principles/

[14] Zenodo explains that the data retention period is the lifetime of the repository, which is currently the lifetime of the host laboratory CERN, defined as at least 20 years

sharing, and reuse. The cost required to make research data compliant with Horizon Europe standards is eligible under the WATERVERSE project, following the obligations of the Grant Agreement. The cost for making the data compliant with FAIR principles can be assumed by the corresponding budget inside the project for each partner involved since the data management following this DMP aligns with the WATERVERSE project's commitments.

3.4 Data security

WATERVERSE project adheres to privacy-by-design and security-by-design principles. Data storage will be managed through the project's shared repository on SharePoint, integrated into Microsoft Teams, and administered by CERTH. SharePoint was selected for its advanced security policies, which include robust encryption mechanisms for data at rest and in transit. As a GDPR-compliant platform, SharePoint will notify CERTH's admin in case of a breach, and periodic backups will be made and updated for data recovery.

All collected data during the lifespan of WATERVERSE project will be handled securely to prevent loss and unauthorized access. Security controls during data collection will be in place including secure channels and encryption algorithms, protecting the data of being tampered or accessed by unauthorised individuals. As far as data storage is concerned, this will be made on data on private servers and secure data management environments through deploying certain protection layers in their premises, granting access only to authorised users on a need-to-know basis. All data will be protected behind robust firewall systems, while partners responsible for processing data will ensure necessary security controls, including frequent backups and recovery systems, integrity checks, and access controls within their infrastructure. Each partner will act as an independent data controller to guarantee the security of the data used for the project's development. Data communication should be done according to their level of confidentiality and the relevant Intellectual Property Rights and Copy Rights in place.

Upon the project's completion, the partner storing the dataset will maintain all responsibilities concerning data recovery and secure storage. If additional safeguards are necessary, the project will use secure cloud services, including the security settings of specific cloud hosting providers. Zenodo will serve as the platform for long-term preservation of open data.

The second iteration of the Data Management Plan (M20) will further detail the adopted security plan, outlining any possible change, corrective actions and mitigation measures to potential risks encountered during the data management.

3.5 Personal Data Protection

One of the key pillars to this Data Management Plan is to safeguard all the processing activities that entail personal data in terms of processing. As personal data processing can pose various types of implications (incl. negative) to the data subjects, it is of primary importance for all the necessary safeguards to be put in place to ensure the necessary level of protection. For this reason, the WATERVERSE project has envisaged a clearly defined framework to abide by all data protection considerations. This section provides an overview of the activities that are planned to include personal data, the relevant measures to assure their adequate level of protection as well as the applicable European and national legislation.

To begin with, all WATERVERSE partners have provided a high-level mapping on the personal data they aim to process, the purpose and the legal ground of the processing, whether it is going to be shared and with whom (incl. recipients in third countries), for how long the data is going to be stored, along with certain policies they have adopted at an organisational level for their protection. The following WPs and tasks have been identified to include personal data processing (Table 8).

WP/Task	Provision of Human participation	Reason for Processing	Lawful basis	Data Responsible
WP1 - Project Management	Daily activities of WATERVERSE project (e.g., consortium meetings, mailing list communication, online activities etc.)	Required for the daily implementation of the WATERVERSE project activities	Art 6, 1 (b)[15] GDPR	CERTH
WP2- WATERVERSE data management approach in water Data Spaces *Task 2.1 Stakeholders identification & engagement* *Task 2.2 Water domain Data Space analysis*	Interviews, questionnaires	Required to set up the stakeholders' forum, facilitate stakeholders' further engagement and identify end-user functional and non-functional requirements	Consent - Art 6, 1 (a) GDPR	KWR, CET, PWN, HST, WBL, SWW, HIDR, KEY
WP3- WATERVERSE Water Data Management Ecosystem *Task 3.3: Cybersecurity solutions for water utilities*	No direct human participation is envisioned. However, data collected from internal sources, honeypots (e.g., system, web, database, application, and security logs), may include personal data, i.e., IP addresses of the attackers. No direct human participation is envisioned. However, from the data collected from external (online)	The use of personal data included in the collected data (i.e., IPs) enables the data controller (CERTH/ITI) to correlate the gathered data to extract more advanced intelligence about cyber threats. The purpose of the data controller is not to collect personal data, but to use this data to	Legitimate interests of the data controller (CERTH/ITI) according to art. 6 (1) (f) GDPR in the sense of Recitals 50 and 49 GDPR, i.e., for scientific research purposes and for the purposes of ensuring network and information security, as these are the	CERTH

[15] Processing is necessary for the performance of a contract to which the data subject is party or to take steps at the request of the data subject prior to entering a contract.

WP/Task	Provision of Human participation	Reason for Processing	Lawful basis	Data Responsible
	sources (e.g., from collaborative and public databases, from social media posts related to cyber security, from forum posts and webpages in the surface and dark web, all CTI related), there is a possibility, due to the nature of the external online sources for personal data to be included (e.g. IP addresses, e-mails, names, surnames, birthplaces etc.)	enrich the collected intelligence.	objectives of the research project.	
WP5 - Setting the WATERVERSE Water Data Management Ecosystem for Water Data Spaces	Training (tutorials) and Pilot Activities at each pilot site	Required to collect structured and semi-structured feedback for the training tutorials. Also, important to perform testing, evaluation and validation focusing on usability tests with system operators in operational environment for the WATERVERSE product	Consent - Art 6, 1 (a) GDPR	PHOEBE, ENG
WP6 - Dissemination, communication and exploitation	Human participation is foreseen apart from the relevant dissemination and communication activities (e.g., newsletter, invitations to	Required for proper and successful fulfilment of the aims and objectives of WP6	Consent - Art 6, 1 (a) GDPR	WE, KWR, PHOBE

WP/Task	Provision of Human participation	Reason for Processing	Lawful basis	Data Responsible
	workshops, conferences, and other networking activities – T6.1, T6.2, T6.3), also to the business model definition (T6.4) for the future exploitation of WATERVERSE solution			

Table 8: Personal data processing in WATERVERSE activities

In all activities that require human as described above, all individuals will be adults, able to provide their informed consent and involved on a voluntary basis, without having to encounter any physical, emotional, social, economic, or legal risk as well as any deceptive activities that would create psychological stress or anxiety under any circumstances. Consortium members will be responsible for their voluntary recruitment and further participation, following the ALLEA European Code of Conduct for Research Integrity[16], namely:

- Reliability
- Honesty
- Respect
- Accountability

Any personal data derived from the participants (e.g., name, surname, e-mail address, image, video, audio, etc.) will abide by the **data minimization principle**, ensuring that in the relevant information sheets produced for each activity there will be well-rounded and sufficient explanation on the reasons for the project to collect the personal data per case and for the purpose of the project. No environmental damage, political or financial adverse consequences and misuse are foreseen as potential outcomes of project's activities. The Consortium will be careful in the way they formulate and publish their research findings to avoid the stigmatization or stereotyping of any of the involved groups.

Informed consent and information sheets including specific information on data management and participants' personal data protection rights will be developed, with details adjusted to the needs of the specific activities and communicated in understandable language to the participants, while an oral briefing will be provided to the research participants prior to research activities and debriefing will always follow. As regards the data collected from online sources (WP3, T3.3) there is always the risk that individuals do not anticipate their data to be used in another way than the one they had uploaded online. For that reason, a data protection notice will be made publicly available in accordance with art. 14 GDPR and with its potentially applicable derogations (art. 14 (5) (b) GDPR[17]), as an effort of enabling the data

[16] https://allea.org/code-of-conduct/

[17] Paragraph 5 (b) of this Article provides for an exemption if such information proves impossible or would involve a disproportionate effort, for processing for archiving purposes in the public interest, scientific or historical research

subjects to be informed about the data processing and to exercise their rights. In this document specific information will be provided concerning the aims and objectives of the WATERVERSE project, the categories of collected and processed data, the purpose of data processing, the contact details of the relevant project representatives, as well as the rights of the data subjects.

The WATERVERSE project implements appropriate technical, organizational and security measures to ensure an appropriate level of protection against the risks arising from processing, such as accidental or unlawful destruction, loss, alteration, unauthorized disclosure, or access (Section 3.4). Personal data considered useful for the project will be either anonymized/pseudonymized and deleted or encrypted and stored in a password-protected database stored in the project SharePoint which will be only accessible to partners that have been given a unique set of usernames and passwords. Appropriate privacy safeguarding measures will be applied by parties involved in the project, limiting the disclosure of personal data, and applying the data minimisation principle to any possible encounter. In case of personal data breaches, project partners will promptly notify their competent national supervisory authorities and affected data subjects, documenting all related information.

All personal data will be properly stored in their pseudonymized/anonymized form for the duration of the project plus five (5) years after the end of the project, to be available for demonstration in case of an inspection or an audit, as long as required to achieve the above purposes of processing, unless a longer retention period is required by law or for the establishment, exercise or defense of legal claims.

Parallel to all the above, the WATERVERSE partners will ensure that no personal data will be shared between the partners unless it has been fully pseudonymized/anonymized prior to the data sharing, or on a need-to-know basis, in accordance with the EU Data Protection Legislation. All partners will determine all operational measures that should be taken prior to any personal data exchange or processing, with certain joint controllership agreements to be in place abiding with the existing rules and regulations of the EU.

3.6 Ethics and legal compliance

It is of highest importance for all projects that involve human participation to abide with all relevant applicable national, EU and international legislation, especially whenever real data inherent to human participants is processed. WATERVERSE project will take into consideration all the relevant EU legislation, to be fully legally ethically compliant when dealing with personal data:

a. The Charter of Fundamental Rights of the European Union[18] 2000/C, 36401with a focus on respect for human dignity, right to the physical and mental integrity of the person and respect for privacy and protection of personal data

b. The EU Ethics and Data protection[19] (05 July 2021) of Horizon Europe

purposes or statistical purposes. In this case, subject to the conditions and safeguards referred to in Article 89(1) GDPR, the controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available.

[18] https://www.europarl.europa.eu/charter/pdf/text_en.pdf

[19] https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf

c. The General Data Protection Regulation[20] (Regulation (EU) 2016/679), and more in particular its provisions around 'personal data' (Art 4 (1) and Recital 26 GDPR), 'the principles relating to the processing of personal data' (Art 5 GDPR), 'the lawfulness of processing' (Art 6 GDPR), 'the rights of the data subject' (Art. 12-23 and 77 GDPR) and also provision around the controller and processor, the security of data as well as special categories of data, also envisioned in GDPR provisions.

d. The Directive 2002/58/EC[21] of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications).

In addition to that all the specific national legislations of the WATERVERSE consortium partner-countries on the protection of personal data, that generally incorporate the GDPR in their territories, have been taken into consideration.

[20] https://eur-lex.europa.eu/eli/reg/2016/679/oj

[21] https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32002L0058

4.0 WATERVERSE DATASETS

Under this section and in the tables below, an initial identification of the datasets along with their characteristics that will be generated/collected and managed by the WATERVERSE partners throughout the course of the project have been performed. This list will be constantly reviewed and updated during the lifespan of the project as soon as the datasets characteristics will be defined in further details.

4.1 WATERVERSE Datasets for User Requirements and Pilot Use Cases (WP2, WP5)

4.1.1 Datasets for Pilot 1 – Netherlands

Dataset Name	PWN_scada_WP5_001
Dataset Type	Existing data, used in the same format
Dataset Category and Access level	Primary data
Data owner/controller – Data Provider(s)	PWN
Time period covered by the Dataset	TBD
Language	English
Expected size of dataset	TBD
Types and formats of Data	csv
Purpose of data generation/re-use	Modelling, visualisation, data exchange tests
WP, Task, Deliverable	WP3 (with link to WP5)
Usefulness of data generation/re-use and data users	Water utility (PWN), research organisation, scientific publications
National / funders / sectorial or departmental procedures for data management	N/A
Other research outputs	Data pipelines to transfer the on-premise data to Azure datalake to deliver project outputs.

Table 9: PWN_scada_WP5_001 Dataset Management

Dataset Name	RWS_API_WP5_001
Dataset Type	Re-use of existing data from the Rijkswaterstaat API, for modelling. https://waterwebservices.rijkswaterstaat.nl
Dataset Category and Access level	Secondary data
Data owner/controller – Data Provider(s)	RWS
Time period covered by the Dataset	TBD
Language	Dutch
Expected size of dataset	TBD
Types and formats of Data	Csv
Purpose of data generation/re-use	modelling, visualisation, data exchange tests
WP, Task, Deliverable	WP3 (with link to WP5)
Usefulness of data generation/re-use and data users	Water utility (PWN), research organisation, scientific publications
National / funders / sectorial or departmental procedures for data management	Rijkswaterstaat is the executive agency of the Ministry of Infrastructure and Water Management and works every day to make the Netherlands safe, liveable and accessible. https://www.rijkswaterstaat.nl/over-ons/onze-organisatie
Other research outputs	Data pipelines to transfer the on-premise data to Azure datalake to deliver project outputs.

Table 10: RWS_API_WP5_001 Dataset Management

Dataset Name	KNMI_API_WP5_001
Dataset Type	Re-use of existing data from the KNMI API, for modeling. https://www.daggegevens.knmi.nl/klimatologie/uurgegevens
Dataset Category and Access level	Secondary data
Data owner/controller – Data Provider(s)	KNMI

Time period covered by the Dataset	TBD
Language	Dutch
Expected size of dataset	TBD
Types and formats of Data	Csv
Purpose of data generation/re-use	modelling, visualisation, data exchange tests
WP, Task, Deliverable	WP3 (with link to WP5)
Usefulness of data generation/re-use and data users	Water utility (PWN), research organisation, scientific publications
National / funders / sectorial or departmental procedures for data management	The Royal Netherlands Meteorological Institute (KNMI) is the national knowledge center and data center for weather, climate and seismology. https://www.rijksoverheid.nl/contact/contactgids/koninklijk-nederlands-meteorologisch-instituut
Other research outputs	Data pipelines to transfer the on-premise data to Azure datalake to deliver project outputs.

Table 11: KNMI_API_WP5_001 Dataset Management

Dataset Name	Aqualarm_API_WP5_001
Dataset Type	Re-use of existing data from the Aqualarm API, for modeling. https://aqualarm.nl/adi/dd-oper/2.0/locations/Lobith/quantities/CLBI/timeseries
Dataset Category and Access level	Secondary data
Data owner/controller – Data Provider(s)	RWS
Time period covered by the Dataset	TBD
Language	Dutch
Expected size of dataset	TBD
Types and formats of Data	Csv

Purpose of data generation/re-use	Modelling, visualisation, data exchange tests
WP, Task, Deliverable	WP3 (with link to WP5)
Usefulness of data generation/re-use and data users	Water utility (PWN), research organisation, scientific publications
National / funders / sectorial or departmental procedures for data management	Rijkswaterstaat is the executive agency of the Ministry of Infrastructure and Water Management and works every day to make the Netherlands safe, liveable and accessible. https://www.rijkswaterstaat.nl/over-ons/onze-organisatie
Other research outputs	Data pipelines to transfer the on-premise data to Azure datalake to deliver project outputs.

Table 12: Aqualarm_API_WP5_001 Dataset Management

4.1.2 Datasets for Pilot 3 – Cyprus

Dataset Name	CY_pilot_dataset_1
Dataset Type	SCADA system timeseries (generated online through the SCADA system)
Dataset Category and Access level	Primary data from system operation, hydraulics (flow, pressure, level) + quality (chlorine residual)
Data owner/controller – Data Provider(s)	Water Board of Lemesos is the owner/controller of the data. PHOEBE will act as data processor, using the data in the framework of the pilot and also making them available through a RESTful API, at the extent required.
Time period covered by the Dataset	Real time, with 5-minute frequency. Available also as historical data for at least 1 year.
Language	English
Expected size of dataset	Depends on the time-range required. Latest values comprise a very small data size. However, if we consider historical data of 1 year, the size is multiplied by the number of sensors and the frequency.
Types and formats of Data	CSV file. Plan to have a JSON-based API in the near future.
Purpose of data generation/re-use	Data comprise real operation data of the water network. Plan to use the data in the DigitalTwin setup of the CY pilot, to facilitate decision making following appropriate analysis.
WP, Task, Deliverable	WP2, WP5

Usefulness of data generation/re-use and data users	Universities for research/innovation purposes. SMEs for the development and deployment of added-value services. Final users can be researchers or software tools developers and Utility operators through the monitoring tools deployed in the control centre.
National / funders / sectorial or departmental procedures for data management	Data are generated and used in compliance with national and EU legislation (i.e., software tools adopt all required security standards, operators and in general people with access to data are authenticated by username/password credentials). The same datasets are used in other EU/national projects for pilot purposes.
Other research outputs	N/A

Table 13: CY_pilot_dataset_1 Dataset Management

Dataset Name	CY_pilot_dataset_2
Dataset Type	Generated data through the DigitalTwin software of PHOEBE
Dataset Category and Access level	Secondary data produced by processing of the primary data.
Data owner/controller – Data Provider(s)	Water Board of Lemesos is the owner/controller of the data. PHOEBE will act as data processor, using the data in the framework of the pilot and making them available through a RESTful API, at the extent required.
Time period covered by the Dataset	Real time, with 5-minute frequency. Available also as historical data for at least 1 year.
Language	English
Expected size of dataset	Depends on the time-range required. Latest values comprise a very small data size. However, if we consider historical data of 1 year, the size is multiplied by the number of sensors and the frequency.
Types and formats of Data	CSV, JSON
Purpose of data generation/re-use	State estimation in network locations without real measurements.
WP, Task, Deliverable	WP5
Usefulness of data generation/re-use and data users	Useful for researchers to develop further algorithms. Useful for system operators to make informed decisions. Useful for 3rd party software tools.
National / funders / sectorial or departmental	Data are generated and used in compliance with national and EU legislation (i.e., software tools adopt all required security standards, operators and in

procedures for data management	general people with access to data are authenticated by username/password credentials). The same datasets are used in other EU/national projects for pilot purposes.
Other research outputs	N/A

Table 14: CY_pilot_dataset_2 Dataset Management

4.1.3 Datasets for Pilot 4 – United Kingdom

Dataset Name	Subset_SCADA_point_hierarchy
Dataset Type	Re use of existing data from SCADA datamart, generated by network sensors
Dataset Category and Access level	Primary data. List of SCADA signals.
Data owner/controller – Data Provider(s)	SWW
Time period covered by the Dataset	n/a signal lists maintained as new signals added / old ones deprecated
Language	English
Expected size of dataset	Small (if restricted to single study catchment). Approx 300 rows.
Types and formats of Data	TBC likely .csv exports from source database via FTP
Purpose of data generation/re-use	To enable human interpretation of what sites/signals the digital and analogue data codes refer to
WP, Task, Deliverable	WP2, WP5
Usefulness of data generation/re-use and data users	Internal SWW users. Subject to data access approval, wider stakeholders such as universities and rivers trusts.
National / funders / sectorial or departmental procedures for data management	Data release outside of SWW must have had appropriate approval from the Data Protection Officer and Senior Management
Other research outputs	N/A

Table 15: Subset SCADA point hierarchy Dataset Management

Dataset Name	Subset_SCADA_analogue
Dataset Type	Re use of existing data from SCADA datamart, generated by network sensors

Dataset Category and Access level	Primary data. SCADA continuous data signals
Data owner/controller – Data Provider(s)	SWW
Time period covered by the Dataset	circa 2019 - ongoing
Language	English
Expected size of dataset	Small-Medium. About 1M rows expanding at 1 row per 15 min per circa 80 signals (if restricted to single study catchment)
Types and formats of Data	TBC likely .csv exports from source database via FTP
Purpose of data generation/re-use	Analogue values will mostly be water level trends to enable UC3 (infiltration) and data validation for UC1
WP, Task, Deliverable	WP2, WP5
Usefulness of data generation/re-use and data users	Internal SWW users. Subject to data access approval, wider stakeholders such as universities and rivers trusts.
National / funders / sectorial or departmental procedures for data management	Data release outside of SWW must have had appropriate approval from the Data Protection Officer and Senior Management
Other research outputs	N/A

Table 16: Subset SCADA analogue Dataset Management

Dataset Name	Subset_SCADA_spill_events
Dataset Type	Re use of existing data from SCADA datamart, generated by network sensors
Dataset Category and Access level	Secondary data. Storm overflow spill events generated by logic acting on digital and analogue signals.
Data owner/controller – Data Provider(s)	SWW
Time period covered by the Dataset	circa 2019 - ongoing
Language	English
Expected size of dataset	Small. 10k rows expanding when sites spills. (If restricted to single study catchment)

Types and formats of Data	TBC likely .csv exports from source database via FTP
Purpose of data generation/re-use	Digital values will mostly be pump on off data to enable UC3 (infiltration) and spill identification/validation for UC1 (e.g. if pumped overflow)
WP, Task, Deliverable	WP2, WP5
Usefulness of data generation/re-use and data users	Internal SWW users. Subject to data access approval, wider stakeholders such as universities and rivers trusts.
National / funders / sectorial or departmental procedures for data management	Data release outside of SWW must have had appropriate approval from the Data Protection Officer and Senior Management
Other research outputs	N/A

Table 17: Subset SCADA spill events Dataset Management

Dataset Name	Subset_SCADA_digital
Dataset Type	Re use of existing data from SCADA datamart, generated by network sensors
Dataset Category and Access level	Primary data. SCADA binary signals.
Data owner/controller – Data Provider(s)	SWW
Time period covered by the Dataset	circa 2019 - ongoing
Language	English
Expected size of dataset	Small-Medium. About 600k rows, expanding over time from aprox 200 signals (new row on state change). If restricted to study catchment
Types and formats of Data	TBC likely .csv exports from source database via FTP
Purpose of data generation/re-use	Identification of period of spills to enable analysis of interactions with water quality data
WP, Task, Deliverable	WP2, WP5
Usefulness of data generation/re-use and data users	Internal SWW users. Subject to data access approval, wider stakeholders such as universities and rivers trusts.
National / funders / sectorial or	Data release outside of SWW must have had appropriate approval from the Data Protection Officer and Senior Management

departmental procedures for data management	
Other research outputs	N/A

Table 18: Subset SCADA digital Dataset Management

Dataset Name	Water_Quality_Monitor
Dataset Type	Re use of existing data from Meteor Cloud
Dataset Category and Access level	Primary data from WQ monitors
Data owner/controller – Data Provider(s)	SWW
Time period covered by the Dataset	circa 2022 - ongoing
Language	English
Expected size of dataset	Small. Temp, cond, ammonia, turbidity, DO from approx 5 WQ monitors. Many removed from river during flood periods.
Types and formats of Data	API (json)
Purpose of data generation/re-use	To monitor water quality in rivers downstream of CSOs as trial in advance of wider rollout under Environment Act
WP, Task, Deliverable	WP2, WP5
Usefulness of data generation/re-use and data users	Internal SWW users. Subject to data access approval, wider stakeholders such as universities and rivers trusts.
National / funders / sectorial or departmental procedures for data management	Data release outside of SWW must have had appropriate approval from the Data Protection Officer and Senior Management
Other research outputs	N/A

Table 19: Water Quality Dataset Management

Dataset Name	EA_rainfall
Dataset Type	Re use of existing EA_rainfall data

Dataset Category and Access level	Primary data from rainfall gauges
Data owner/controller – Data Provider(s)	EA
Time period covered by the Dataset	Ongoing
Language	English
Expected size of dataset	Small (only single local rain gauge)
Types and formats of Data	API (json)
Purpose of data generation/re-use	To allow linkages between rainfall, CSO spillage and water quality. May also be possible to provide rain radar data.
WP, Task, Deliverable	WP2, WP5
Usefulness of data generation/re-use and data users	Internal SWW users, wider stakeholders
National / funders / sectorial or departmental procedures for data management	EA API T+Cs
Other research outputs	N/A

Table 20: EA Rainfall Dataset Management

Dataset Name	EA_river_level
Dataset Type	Re use of existing EA river data
Dataset Category and Access level	Primary data for river gauges
Data owner/controller – Data Provider(s)	EA
Time period covered by the Dataset	Ongoing
Language	English
Expected size of dataset	Small (single gauging station)

Types and formats of Data	API (json)
Purpose of data generation/re-use	To allow linkages between river level / increased rural runoff and water quality
WP, Task, Deliverable	WP2, WP5
Usefulness of data generation/re-use and data users	Internal SWW users, wider stakeholders
National / funders / sectorial or departmental procedures for data management	EA API T+Cs
Other research outputs	N/A

Table 21: EA River Level Dataset Management

4.1.4 Datasets for Pilot 5 – Spain

Dataset Name	SP01_DM_SmartMetering_Flow
Dataset Type	Reused data from smart metering platform will be used to generate KPIs
Dataset Category and Access level	Smart metering flow measures of customers and several points of the network
Data owner/controller – Data Provider(s)	Hidralia is the data owner/controller but the data should be provided by another company of its business group
Time period covered by the Dataset	Time frequency from 15 minutes to 1 hour. At least 1 year of data should be provided
Language	This dataset only contains IDs and numbers so there isn't any associated language to the dataset
Expected size of dataset	Depends on the number of customers to be covered by the pilot and the time period
Types and formats of Data	Float
Purpose of data generation/re-use	It will be used to generate KPIs regarding the efficiency of the network and some others regarding the customers behaviour
WP, Task, Deliverable	WP2, WP5
Usefulness of data generation/re-use and data users	This information will be used internally in order to calculate the KPIs.

National / funders / sectorial or departmental procedures for data management	Data should be used in compliance with national and EU legislation. Concretely it should be in compliance with GDPR and Esquema Nacional de Seguridad from Spanish government.
Other research outputs	N/A

Table 22: SP01_DM_SmartMetering_Flow Dataset Management

Dataset Name	SP02_CF_ManualMetering_Customers
Dataset Type	Reused data from customer platform will be used to generate KPIs
Dataset Category and Access level	Manual metering flow measures of customers not covered by smart metering platform
Data owner/controller – Data Provider(s)	Hidralia is the data owner/controller, but the data should be provided by another company of its business group
Time period covered by the Dataset	Time frequency 1 month. At least 1 year of data should be provided
Language	This dataset only contains IDs and numbers so there isn't any associated language to the dataset
Expected size of dataset	Depends on the number of customers to be covered by the pilot and the time period
Types and formats of Data	Float
Purpose of data generation/re-use	It will be used to generate KPIs regarding the efficiency of the network and some others regarding the customers behaviour
WP, Task, Deliverable	WP2, WP5
Usefulness of data generation/re-use and data users	This information will be used internally in order to calculate the KPIs.
National / funders / sectorial or departmental procedures for data management	Data should be used in compliance with national and EU legislation. Concretely it should be in compliance with GDPR and Esquema Nacional de Seguridad from Spanish government.
Other research outputs	N/A

Table 23: SP02_CF_ManualMetering_Customers Dataset Management

Dataset Name	SP03_GA_SmartMetering_Network
Dataset Type	Reused data from SCADA will be used to generate KPIs
Dataset Category and Access level	Flow, pressure, state and levels measures from several points of the network
Data owner/controller – Data Provider(s)	Hidralia is the data owner/controller, but the data should be provided by another company of its business group
Time period covered by the Dataset	5-minute frequency. At least 1 year of data should be provided
Language	This dataset only contains IDs and numbers so there isn't any associated language to the dataset
Expected size of dataset	Depends on the number of customers to be covered by the pilot and the time period
Types and formats of Data	Float
Purpose of data generation/re-use	It will be used to generate KPIs regarding the efficiency of the network and some others regarding the customers behaviour
WP, Task, Deliverable	WP2, WP5
Usefulness of data generation/re-use and data users	This information will be used internally in order to calculate the KPIs.
National / funders / sectorial or departmental procedures for data management	Data should be used in compliance with national and EU legislation. Concretely it should be in compliance with GDPR and Esquema Nacional de Seguridad from Spanish government.
Other research outputs	N/A

Table 24: SP03_GA_SmartMetering_Network Dataset Management

Dataset Name	SP04_iZeus_Energy
Dataset Type	Reused data from energy platform will be used to generate KPIs
Dataset Category and Access level	energy generation and consumption from several points of the network
Data owner/controller – Data Provider(s)	Hidralia is the data owner/controller, but the data should be provided by another company of its business group
Time period covered by the Dataset	TBD

Language	This dataset only contains IDs and numbers so there isn't any associated language to the dataset
Expected size of dataset	Depends on the number of customers to be covered by the pilot and the time period
Types and formats of Data	Float
Purpose of data generation/re-use	It will be used to generate KPIs regarding the efficiency of the network and some others regarding the customers behaviour
WP, Task, Deliverable	WP2, WP5
Usefulness of data generation/re-use and data users	This information will be used internally in order to calculate the KPIs.
National / funders / sectorial or departmental procedures for data management	Data should be used in compliance with national and EU legislation. Concretely it should be in compliance with GDPR and Esquema Nacional de Seguridad from Spanish government.
Other research outputs	N/A

Table 25: SP04_iZeus_Energy Dataset Management

Dataset Name	SP05_KPIs
Dataset Type	Generated within the pilot
Dataset Category and Access level	KPIs regarding efficiency and other indicators of performance of citizens behaviour
Data owner/controller – Data Provider(s)	Hidralia is the data owner/controller, but the data should be provided by another company of its business group
Time period covered by the Dataset	TBD
Language	This dataset only contains IDs and numbers so there isn't any associated language to the dataset
Expected size of dataset	TBD
Types and formats of Data	TBD
Purpose of data generation/re-use	Generated data will be used to increase the vision of Management performance and provide useful information to third parties about citizens behaviour

WP, Task, Deliverable	WP5
Usefulness of data generation/re-use and data users	Universities for research/innovation purposes. SMEs for the development and deployment of added-value services. Final users can be researchers or software tools developers and Utility operators through the monitoring tools deployed in the control centre.
National / funders / sectorial or departmental procedures for data management	Data should be used in compliance with national and EU legislation. Concretely it should be in compliance with GDPR and Esquema Nacional de Seguridad from Spanish government.
Other research outputs	N/A

Table 26: SP05_KPIs Dataset Management

4.1.5 Datasets for Pilot 6 – Finland

Dataset Name	dataset_FI_1_devicesignal
Dataset Type	existing data in pilot customer
Dataset Category and Access level	primary data, private, json signal
Data owner/controller – Data Provider(s)	Water utility (tbc)
Time period covered by the Dataset	Real time signal
Language	English, Finnish
Expected size of dataset	TBD
Types and formats of Data	json, raw signal
Purpose of data generation/re-use	For converting data to NGSI-LD and pilot with existing tools
WP, Task, Deliverable	WP5 (and WP3)
Usefulness of data generation/re-use and data users	Water utilities
National / funders / sectorial or departmental	N/A

procedures for data management	
Other research outputs	N/A

Table 27: dataset_FI_1_devicesignal Dataset Management

Dataset Name	dataset_FI_2_networkobject
Dataset Type	existing data in pilot customer
Dataset Category and Access level	primary data, private, json signal
Data owner/controller – Data Provider(s)	Water utility (tbc)
Time period covered by the Dataset	About 25 year -now, real time
Language	English, Finnish
Expected size of dataset	TBD
Types and formats of Data	json
Purpose of data generation/re-use	For converting data to NGSI-LD and pilot with existing tools
WP, Task, Deliverable	WP5 (and WP3)
Usefulness of data generation/re-use and data users	Water utilities
National / funders / sectorial or departmental procedures for data management	N/A
Other research outputs	N/A

Table 28: dataset_FI_2_networkobject Dataset Management

4.2 WATERVERSE Datasets for Data Management ecosystem and Cyber Security Solutions (WP3)

Dataset Name	AI_DVR_WP3_NL_use_case
Dataset Type	Generate data through the functioning of the AI-based data validation tool (RO#6)
Dataset Category and Access level	Primary data
Data owner/controller – Data Provider(s)	KWR, PWN
Time period covered by the Dataset	TBD
Language	English
Expected size of dataset	TBD
Types and formats of Data	TBD
Purpose of data generation/re-use	Validation of raw sensor data and screening for anomalies
WP, Task, Deliverable	WP3
Usefulness of data generation/re-use and data users	water utility (PWN), research organisation (KWR), scientific publications
National / funders / sectorial or departmental procedures for data management	N/A
Other research outputs	Research output includes a tool as a software/middleware included relevant workflows and scripts for execution within the WDME.

Table 29: AI_DVR_WP3_NL_use_case Dataset Management

Dataset Name	AI_DVR_WP3_DE_use_case
Dataset Type	Generate data through the functioning of the AI-based data validation tool (RO#6)
Dataset Category and Access level	Primary data

Data owner/controller – Data Provider(s)	HST, FIWARE
Time period covered by the Dataset	TBD
Language	English, German
Expected size of dataset	TBD
Types and formats of Data	TBD
Purpose of data generation/re-use	Validation of raw sensor data and screening for anomalies
WP, Task, Deliverable	WP3
Usefulness of data generation/re-use and data users	research organisation (HST, FIWARE), municipality (Etteln)
National / funders / sectorial or departmental procedures for data management	N/A
Other research outputs	Research output includes a tool as a software/middleware included relevant workflows and scripts for execution within the WDME.

Table 30: AI_DVR_WP3_DE_use_case Dataset Management

Dataset Name	AI_DVR_WP3_CY_use_case
Dataset Type	Generate data through the functioning of the AI-based data validation tool (RO#6)
Dataset Category and Access level	Primary data
Data owner/controller – Data Provider(s)	KWR, WBL, PHOEBE
Time period covered by the Dataset	TBD
Language	English
Expected size of dataset	TBD

Types and formats of Data	TBD
Purpose of data generation/re-use	Validation of raw sensor data and screening for anomalies
WP, Task, Deliverable	WP3
Usefulness of data generation/re-use and data users	water utility (WBL), research organisation (KWR), SMEs (PHOEBE), scientific publications
National / funders / sectorial or departmental procedures for data management	N/A
Other research outputs	Research output includes a tool as a software/middleware included relevant workflows and scripts for execution within the WDME.

Table 31: AI_DVR_WP3_CY_use_case Dataset Management

Dataset Name	AI_DVR_WP3_UK_use_case
Dataset Type	Generate data through the functioning of the AI-based data validation tool (RO#6)
Dataset Category and Access level	Primary data
Data owner/controller – Data Provider(s)	KWR, SWW, UNIEXE
Time period covered by the Dataset	TBD
Language	English
Expected size of dataset	TBD
Types and formats of Data	TBD
Purpose of data generation/re-use	Validation of raw sensor data and screening for anomalies
WP, Task, Deliverable	WP3
Usefulness of data generation/re-use and data users	water utility (SWW), research organisation (KWR, UNIEXE), scientific publications

National / funders / sectorial or departmental procedures for data management	N/A
Other research outputs	Research output includes a tool as a software/middleware included relevant workflows and scripts for execution within the WDME.

Table 32: AI_DVR_WP3_UK_use_case Dataset Management

Dataset Name	AI_DVR_WP3_ES_use_case
Dataset Type	Generate data through the functioning of the AI-based data validation tool (RO#6)
Dataset Category and Access level	Primary data
Data owner/controller – Data Provider(s)	KWW, HIDR, CET
Time period covered by the Dataset	TBD
Language	English
Expected size of dataset	TBD
Types and formats of Data	TBD
Purpose of data generation/re-use	Validation of raw sensor data and screening for anomalies
WP, Task, Deliverable	WP3
Usefulness of data generation/re-use and data users	water utility (HIDR), research organisation (KWR, CET), scientific publications
National / funders / sectorial or departmental procedures for data management	N/A
Other research outputs	Research output includes a tool as a software/middleware included relevant workflows and scripts for execution within the WDME.

Table 33: AI_DVR_WP3_ES_use_case Dataset Management

Dataset Name	CTI_Data
Dataset Type	We will re-use data collected from threat related databases and from our own honeypots. The data will be used to populate the threat intelligence module and to train our ML-models
Dataset Category and Access level	Primary and secondary data. Both publicly and not publicly available data. CTI related data from vulnerability databases, CERT feeds, databases with Proof-of-Concept exploits, social media, forums, and relevant web pages from the Surface and the Dark Web and from honeypots.
Data owner/ controller – Data Provider(s)	CERTH / PWN, HIDR, KEY, WBL, HST, SWW
Time period covered by the Dataset	TBD
Language	English and machine code
Expected size of dataset	TBD
Types and formats of Data	STIX, SQL, JAVA, Twitter, MySQL, JSON, TBD
Purpose of data generation/re-use	We will re-use data collected from threat related databases and from our own honeypots. The data will be used to populate the threat intelligence module and to train our ML-models. It is mandatory for the activities of T3.3
WP, Task, Deliverable	WP3 (T3.3)
Usefulness of data generation/re-use and data users	Research organisations, scientific community, Network of water utilities
National / funders / sectorial or departmental procedures for data management	N/A
Other research outputs	N/A

Table 34: CTI_Data Dataset Management

Dataset Name	System_and_network_logs
Dataset Type	We will use data from the end-users, and we will re-use data from our honeypots. The data will be used to train our models

Dataset Category and Access level	Primary and secondary data. Both publicly and not publicly available data. Network and system logs from the end-users and from our own honeypots.
Data owner/ controller – Data Provider(s)	End-users and CERTH / PWN, HIDR, KEY, WBL, HST, SWW
Time period covered by the Dataset	TBD
Language	English and machine code
Expected size of dataset	TBD
Types and formats of Data	network and system logs
Purpose of data generation/re-use	We will re-use data collected from threat related databases and from our own honeypots. The data will be used to populate the threat intelligence module and to train our ML-models. It is mandatory for the activities of T3.3
WP, Task, Deliverable	WP3 (T3.3)
Usefulness of data generation/re-use and data users	Research organisations, scientific community, Network of water utilities
National / funders / sectorial or departmental procedures for data management	N/A
Other research outputs	N/A

Table 35: System_and_network_logs Dataset Management

Dataset Name	Detected threats and anomalies dataset
Dataset Type	We will use data from the end-users, and we will re-use data from our honeypots. The data will be used to train our models.
Dataset Category and Access level	Primary and secondary data. Both publicly and not publicly available data. Network and system logs from the end-users and from our own honeypots.
Data owner/controller – Data Provider(s)	End-users and CERTH / PWN, HIDR, KEY, WBL, HST, SWW
Time period covered by the Dataset	TBD
Language	English and machine code

Expected size of dataset	TBD
Types and formats of Data	STIX, SQL, JAVA, Twitter, MySQL, JSON, TBD
Purpose of data generation/re-use	We will re-use data collected from threat related databases and from our own honeypots. The data will be used to populate the threat intelligence module and to train our ML-models. It is mandatory for the activities of T3.3
WP, Task, Deliverable	WP3 (T3.3)
Usefulness of data generation/re-use and data users	Research organisations, scientific community, Network of water utilities
National / funders / sectorial or departmental procedures for data management	N/A
Other research outputs	N/A

Table 36: Detected threats and anomalies Dataset Management

4.3 WATERVERSE Datasets: Project Management, User Requirements, Stakeholders Network, Communication, Dissemination and Exploitation (WP1, WP2, WP6)

Dataset Name	Project Communication mailing list
Dataset Type	Information on team members of WATERVERSE partners concerning the mailing lists that should participate
Dataset Category and Access level	Primary data
Data owner/controller – Data Provider(s)	CERTH
Time period covered by the Dataset	Project Lifespan
Language	English
Expected size of dataset	Less than 100 KBs

Types and formats of Data	*xlsx, *csv
Purpose of data generation/re-use	To create the internal e-mail lists that will be used for facilitating communication among partners and exchanging information.
WP, Task, Deliverable	WP1
Usefulness of data generation/re-use and data users	Consortium partners

Table 37: Project Communication mailing list Dataset Management

Dataset Name	UR_Specification
Dataset Type	Survey/Questionnaire data
Dataset Category and Access level	Primary data
Data owner/controller – Data Provider(s)	CET
Time period covered by the Dataset	Project Lifespan
Language	English + TBD
Expected size of dataset	TBD
Types and formats of Data	*xlsx, *csv, *docx, *pdf, survey in google forms
Purpose of data generation/re-use	Definition of the functional and non-functional requirements of the WATERVERSE system.
WP, Task, Deliverable	WP2
Usefulness of data generation/re-use and data users	Consortium partners

Table 38: UR_Specification Dataset Management

Dataset Name	WATERVERSE_Stakeholder_Group_Contact_List
Dataset Type	Information on participants to WATERVERSE Stakeholders' group
Dataset Category and Access level	Primary data
Data owner/controller – Data Provider(s)	KWR

Time period covered by the Dataset	Project Lifespan
Language	English
Expected size of dataset	TBD
Types and formats of Data	*xlsx
Purpose of data generation/re-use	Identify and engage relevant stakeholders to end user requirements specification, pilot participation as well as promotion of relevant communication, dissemination and exploitation activities.
WP, Task, Deliverable	WP2
Usefulness of data generation/re-use and data users	Consortium partners, Research Community, Water Utilities

Table 39: WATERVERSE_Stakeholder_Group_Contact_List Dataset Management

Dataset Name	WATERVERSE_Newsletter_Subscribers
Dataset Type	Information on newsletter subscribers (through online application via the project website)
Dataset Category and Access level	Primary data
Data owner/controller – Data Provider(s)	WE
Time period covered by the Dataset	Project Lifespan
Language	English
Expected size of dataset	TBD
Types and formats of Data	*xlsx
Purpose of data generation/re-use	Communication and dissemination activities
WP, Task, Deliverable	WP6
Usefulness of data generation/re-use and data users	Consortium partners, Research and Scientific Community, Water Utilities

Table 40: WATERVERSE_Newsletter_Subscribers Dataset Management

Dataset Name	WATERVERSE_ Events_and_Publications
Dataset Type	Information about events/conferences/publication regarding the WATERVERSE project
Dataset Category and Access level	Primary data
Data owner/controller – Data Provider(s)	WE
Time period covered by the Dataset	Project Lifespan
Language	English
Expected size of dataset	TBD
Types and formats of Data	*xlsx
Purpose of data generation/re-use	Communication and dissemination activities
WP, Task, Deliverable	WP6
Usefulness of data generation/re-use and data users	Consortium partners, Research and Scientific Community, Water Utilities

Table 41: WATERVERSE_ Events_and_Publications Dataset Management

5.0 Conclusion and Future Outlook

This deliverable outlined the WATERVERSE's Self-Assessment Plan (SAP) per Scientific Objective (SO) as well as the Data Management Plan (DMP). Since the project is at an early stage, therefore these plans are susceptible to changes and modifications as WATERVERSE will be evolving.

However, the current status of SAP describes the way that each SO will be monitored and evaluated in order to achieve its goals and outreach the Research Outputs (ROs). The evaluation strategies and the respective KPIs per SO have been introduced. This will lead to a solid monitoring of these activities which could facilitate to receive timely the appropriate mitigation actions to face potential risks and mishaps that may arise.

Furthermore, in this deliverable, the first version of the DMP is introduced. Its aim is to describe in detail the datasets that will be collected/generated/processed during the project as well as to set up the guidelines for the accessibility, interoperability, reusability etc. of these datasets. Currently 33 datasets have been described, the majority of them (20) are related to the datasets provided by the pilots and will be used during the demonstrations of the pilots' use cases. Eight of the datasets concern data that are generated via processes for data management and for the application of cyber security solutions. The rest of them are related to the project management, user requirements specification and communication and dissemination activities.

The DMP plan will be continuously updated and enhanced with any new datasets that needed to be collected by the partners, thus making this report a living document, with its updated versions to be delivered both in the relevant WP4 deliverables and the second version of this deliverable, namely D1.3 Self-assessment and data management plan v2 (M20), to reflect any changes in data management policies or strategies that may occur during the project's lifecycle.

ANNEX A: DATA MANAGEMENT PLAN TEMPLATE

DATASETS Information	
Dataset Name	
Please provide a meaningful name so that we can refer to it unambiguously in the future e.g. Dataset_<WPno>_<serial no.of dataset>_<dataset title>	
Dataset Type	
Please state whether you will generate the data or you will re-use existing data. In case of re-use please describe how you will re-use this data and that is the origin. State the reason if re-use of any existing data has been considered but discarded	
Dataset Category and Access level	
(e.g., project management data, primary data, secondary data, synthetic data, publicly available datasets, etc. You can also add Keywords or phrases describing the subjects or content of the data)	
Data owner/controller – Data Provider(s)	
Please add the names of the partners that will provide the dataset, if this is different from the partner responsible per se, as well as of the data owners of this data set, if it is different from the provider	
Time period covered by the Dataset	
Please state the time period that could be covered by this dataset, if applicable	
Language	
Please state the languages used in the dataset	
Expected size of dataset	
Please state the volume of the data if possible	
Types and formats of Data	
Please outline the formats of the data e.g., doc, docx, xls, pdf, jpg, etc.	
Purpose of data generation/re-use	
Please state the purpose of the data generation/re-use and its relation to the objectives of the project	
WP/Task/Deliverable	

Please state the relevant WP, Task, Deliverable these data are linked to the data generated/re-used	
Usefulness of data generation/re-use and data users	
Please state the end users for whom the data generated/re-used can be useful, e.g. university, research organization, SME's, scientific publication, and who will be the final users of this data	
National / funders / sectorial or departmental procedures for data management	
Please state whether any other national / funders / sectorial or departmental procedures for data management that you make use of. If yes, please describe them	
Security	
Please indicate the general provisions are or will be in place for data security	
What measures will be implemented to avoid loss of data?	
What data recovery measures will you implement in case data is lost?	
Other research outputs	
Apart from the data, are there any other research outputs (e.g. software, workflows, protocols) that need to have a consistent management plan on your behalf?	
Personal Data Information (if applicable)	
Type of personal data collected/processed	
Please state whether you will collect/use personal data e.g., names, contact details, date of birth, etc. Please also refer if this info will include special	
Purpose of personal data collection/processing	
Please state the purpose of the personal data collection/processing and its relation to the objectives of the project. Please also explain why this purpose cannot be achieved with less or no personal data	
Legal basis	
Under which legal ground as outlined by Art. 6/ Art. 9 GDPR do you process the personal information	
Data owner/responsible	
Please add the names of the partners that will be responsible for this data	
WP/Task/Deliverable	

Please state the relevant WP, Task, Deliverable these personal data are linked with	
Dataset Category and access level	
Please state whether these personal data stem from publicly available datasets, if they are generated during the project and also the access level inside the project, e.g. confidential, restricted to specific partners?	
Way/Mode of collection	
Please state how you aim to collect these data e.g., interviews, questionnaire, other activities such as crawling, open source data sets etc	
Storage format	
Please state how you aim store these data e.g., hardcopies, digitally, etc	
Sharing mechanisms	
Please state if the personal data need to be transmitted between different entities, and if yes please describe the workflow	
Organizational policy on protection of personal data	
Please here specify the measures you have in place to securely obtain, protect and store specifically the personal data	
Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?	
Designation of DPO	
Has your organization designated a DPO? If yes please share the details, if not, please state the reason	
FAIR PRINCIPLES identification	
Standards and Metadata Information	
Will data be identified by a persistent identifier?	
Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? If yes, please outline any specific standard identification mechanism you use	
If there are no standards in your discipline describe what type of metadata will be created and how	
Please state the approach you use towards key-words provided that optimise possibilities for re-use	

Will metadata be offered in such a way that it can be harvested and indexed?	
Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?	
Will metadata be guaranteed to remain available after data is no longer available?	
Archiving and preservation (including storage and backup)	
Please provide us information on the repository you will deposit the data, along with the appropriate arrangements, for both openly available and restricted data (e.g. Zenodo, Github, Githab)	
Does the repository ensure that the data is assigned an identifier?	
Will the repository resolve the identifier to a digital object?	
Will the data be safely stored in trusted repositories for long term preservation and curation?	
Data Sharing and accessibility	
Will all data be made openly available? If not please indicate the reason why certain datasets cannot be shared (e.g. legal and contractual reasons etc.)	
When will you make the data available? Please inform in case of licensing or embargo periods.	
With whom will you share the data and under what conditions?	
How will the data be made accessible? (e.g. through a free and standardized access protocol, etc.)	
What methods or software tools are needed to access the Data?	
If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?	
How will the identity of the person accessing the data be ascertained?	
How long will the data remain available and findable?	
Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?	
Is there a need for a Data access committee?	

Data Interoperability	
Will the Data produced in the project be interoperable?	
What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines?	
Will you follow community-endorsed interoperability best practices? If yes which?	
In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?	
Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?	
Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?	
Data Re-use	
Will your data be made freely available in the public domain to permit the widest re-use possible? If yes how will they be licenced?	
Will the data produced in the project be useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.	
How long is it intended that the data remains re-usable?	
Will the provenance of the data be thoroughly documented using the appropriate standards?	
Do you have in place certain data quality assurance processes? If yes, which ones?	
Allocation and Resources	
Please indicate the relevant costs be for making data or other research outputs FAIR in your project	
How will these be covered?	
Who will be responsible for data management in your organisation?	
How will long term preservation be ensured? Have the relevant needed resources been discussed?	

WATERVERSE

Co-funded by
the European Union