ULTIMATE

WATER SMART INDUSTRIAL SYMBIOSIS

D7.3.

Data Management Plan

Authors: Joep van den Broeke (KWR), Christos Makropoulos (NTUA, KWR), Aitor Corchero Rodriguez (EUT)

Date: 07/12/2020

Technical References

Project Acronym	ULTIMATE
Project Title	ULTIMATE: indUstry water-utiLiTy symbIosis for a sMarter wATer society
Project Coordinator	KWR
Project Duration	48 months

Deliverable No.	D7.3.
Dissemination level [1]	PU
Work Package	WP7
Task	7.4
Lead beneficiary	KWR
Contributing beneficiary(ies)	KWR, NTUA, EUT
Author(s)	Joep van den Broeke, Christos Makropoulos, Aitor Corchero Rodriguez
Quality Assurance	Francesco Fatone (UNIVPM), Marika Fantoni (UNIVPM), Anne Kleyböcker (KWB)
Due date of deliverable	30/11/2020
Actual submission date	08/12/2020

[1] PU = Public
   PP = Restricted to other programme participants (including the Commission Services)
   RE = Restricted to a group specified by the consortium (including the Commission Services)
   CO = Confidential, only for members of the consortium (including the Commission Services)

Document history

V	Date	Author(s) /Reviewer(s) (Beneficiary)	Description
0.1	20/07/2020	Joep van den Broeke (KWR)	First Draft
0.2	17/09/2020	Joep van den Broeke (KWR), Christos Makropoulos (NTNU, KWR)	Second Draft
0.3	19/11/2020	Joep van den Broeke (KWR), Aitor Corchero Rodriguez (EUT)	Third Draft
0.3.1	23/11/2020	Francesco Fatone, Marika Fantoni (UNIVPM)	First internal revision
0.4	27/11/2020	Joep van den Broeke (KWR)	UNIVPM Reviewer comments addressed
0.5	07/12/2020	Anne Kleyböcker (KWB)	Second internal revision
1.0	07/12/2020	Joep van den Broeke (KWR)	KWB reviewer comments addressed, Final version

Summary

The purpose of the Data Management Plan (DMP) to consider the many aspects of data management, data and metadata generation, data preservation- maintenance- and analysis, whilst ensuring that data is well managed at present and prepared for preservation in the future. This Data Management Plan is compiled according to the Guidelines on FAIR Data Management in H2020, and the Guidelines to the Rules on the Open Access to Scientific Publications and Open Data Access to Research Data in H2020.

The DMP establishes the development, execution and supervision of plans, policies, programmes and practices that control, protect, deliver and enhance the value of data and information assets. Based on this definition, the following process and procedures for data management are defined:

- **Data governance, such as standards management and guidelines**
- **Data architecture, analysis and design including modelling.**
- **Data maintenance, administration, and data mapping across building blocks and solutions modules.**
- **Data security management including data access, archiving, privacy and security.**
- **Data quality management including query management, data integrity, data quality and quality assurance.**
- **Reference and master data management including data integration, external data transfer, master data management, reference data.**
- **Document, record and content management.**
- **Metadata management, i.e., metadata definition, discovery, publishing, metrics and standardization.**

The DMP provides an analysis of these elements considering all datasets to be elaborated and integrated in ULTIMATE. The DMP corresponds to the current status of knowledge within the consortium about the data that will be generated, collected, stored and processed. As such, the Data Management Plan (DMP) is not a fixed document but will evolve as the project develops. More specifically, this deliverable presents the first version of the DMP for the ULTIMATE project. The DMP will be updated at least in Month 30 (D7.6) and Month 48 (D7.7) of the project.

Disclaimer

This publication reflects only the author's views and the European Union is not liable for any use that may be made of the information contained therein.

Table of Contents

List of Abbreviations

CoP: Community of Practice

CS: Case Study

CTG: Cross Cutting Technology Group

DMP: Data Management Plan

DOI: Digital Object Identifiers

FAIR: Findability, accessibility, interoperability, and reusability

FOAF: Friend of a Friend

GDPR: General Data Protection Regulation

ICT: Information and Communications Technology

KPI: Key Performance Indicator

OGC: Open Geospatial Consortium

PM: Person Month

SNS: Social Networking Sites

TEB: Technology Evidence Base

URL: Uniform Resource Locator

UTF: Unicode Transformation Format

WP: Work Package

WSIS: Water Smart Industrial Symbiosis

1. Introduction

1.1. Purpose of this document

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the ULTIMATE project with regards to all the datasets and information that will be generated by the project. The DMP is not a fixed document but evolves during the lifespan of the project; in fact, it functions as a dynamic document of agreements. The DMP should address the points presented below on a dataset by dataset basis and should reflect the current status of reflection within the consortium about the data that will be generated, collected, stored and processed.

In principle, publicly funded research data are a public good, produced for the public interest that should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property. On this basis, the DMP intends to help researchers consider at an early stage, when research is being designed and planned, how data will be managed during the research process and shared afterwards with the wider research community. Moreover, the DMP is also a prerequisite to participate under the Open Research Data Pilot

The benefits of a well-designed DMP not only concern the way data are treated but also the successful outcome of the project itself. A properly planned DMP guides the researchers first to think what to do with the data and then how to collect, store and process them, etc. Furthermore, a planning in data treatment is important for addressing timely security, privacy and ethical aspects. Considering the data lifecycle, the research data are kept in track in cases of possible staff or other changes. The DMP can also increase preparedness for possible data requests. In short, planned activities, such as implementation of well-designed DMP, stand a better chance of meeting their goals than unplanned ones.

The process of planning is also a process of communication, increasingly important in a multi-partner research. The characteristics of collaboration should be accordingly harmonised among project partners from different organisations or different countries. The DMP also provides an ideal opportunity to engender best practice with regards to e.g. file formats, metadata standards, storage and risk management practices. Thus ULTIMATE is aligned to greater longevity and sustainability of data and higher quality standards.

Ultimately, the goal of ULTIMATE is not only engaging researchers in conversations with service provides, but also to connect to citizens and other stakeholders through living labs, communities of practice and the Technology Evidence Base (TEB) and

digital marketplace. In this context, the DMP is a document that not only organises data sharing, editing and monitoring among those contributing to the project. But it also governs data and information privacy and ethical aspects of data and information storage and sharing. This DMP complements the Consortium Agreement which regulates also restrictions on data sharing and Intellectual Property Management.

This DMP is modelled on the Horizon 2020 FAIR Data Management Plan (DMP) template, included as Annex 1 in the Guidelines on FAIR Data Management in Horizon 2020 (Version 3.0, 26 July 2016).

1.2. Readers Guide

This deliverable is intended for use internally in the project only and provides guidance on data management to the project partners and participants. It is particularly relevant for partners responsible for data collection and pilots. The Data Management Plan (DMP) is not a fixed document but will evolve as the project develops and as new procedures etc. are added or existing ones are changed. More specifically, this deliverable presents the first version of the DMP for the ULTIMATE project. The DMP will be updated at least in Month 30 (D7.6) and Month 48 (D7.7) of the project.

2. Data Summary

2.1. The purpose of the data collection/generation

ULTIMATE will act as a catalyst for 'Water Smart Industrial Symbiosis' (WSIS) in which water/wastewater plays a key role both as a reusable resource but also as a vector for energy and materials to be extracted, treated, stored and reused within a dynamic socio-economic and business oriented industrial ecosystem. ULTIMATE adopts an evidence-based approach anchored on 9 large-scale demonstrations across Europe and south east Mediterranean relevant to the agro-food processing, beverages, heavy chemical/petrochemical and biotech industries (Figure 2.1).
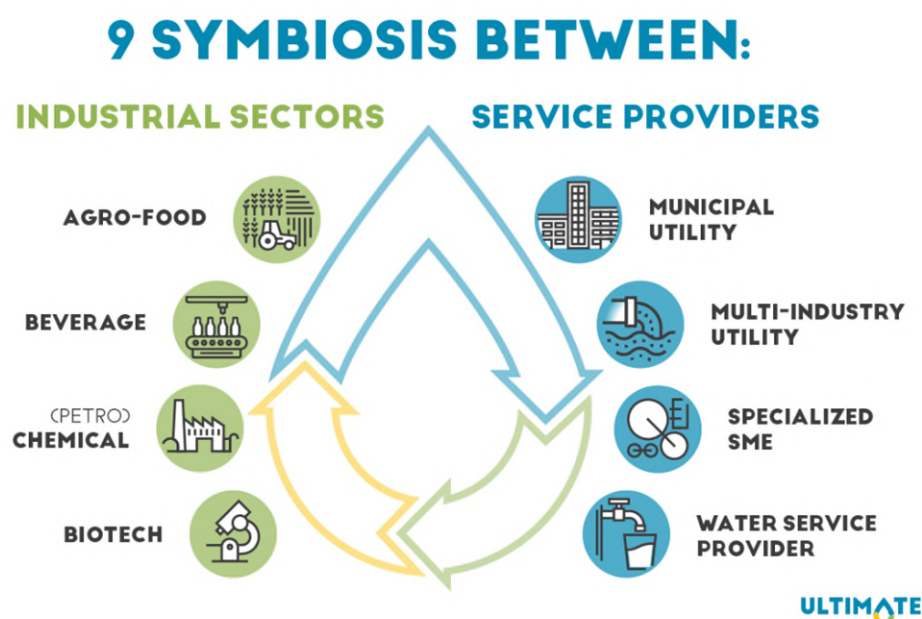

Figure 2.1: Overview of the different industries and services providers collaborating in ULTIMATE WSIS.

In order to accomplish this catalysing role, it is necessary to develop practical approaches, tools and partnerships as well as policy and legal framework conditions, which will accelerate acceptance of these technologies and will broaden the market reach and uptake. The purpose of the data collection and data generation in ULTIMATE is to provide input in the development of the WSIS approaches, the underpinning of the evidence-based approach, and to support building partnerships between business (incl. industrial and technological ecosystems), water service providers, regulators and policy makers. Furthermore, it supports the valorisation of the WSIS solutions by identifying and linking to emerging business opportunities.

2.2. Relation to the objectives of the project

ULTIMATE builds on nine **Large Scale Demonstration Cases** where next generation circular systems and services (in which water plays a key role) are either already deployed at full scale or are to be deployed in the near future, demonstrating the replication potential of the approaches in practice. These real-world demonstrations allow the creation and dissemination of a comprehensive Technology Evidence Base (TEB) (task 1.1, WP1) of the proposed solutions' effectiveness (WP1, WP2), performance (WP1), enabling conditions (WP1, WP2, WP3, WP4) and opportunities for business development (WP5). The information in the TEB will be enhanced by the outcomes of the semantic modelling of industrial processes, simulations, stress-testing, process optimization and redesigning results both at the individual and whole system levels performed in WP2. Furthermore, results from WP2's life cycle assessment, risk analysis, and cost and benefits analysis of WSIS solutions will provide insights into potential for full-scale replicability of the solutions, and will feed into the TEB as well as WP5. Crucially, ULTIMATE promote active stakeholder engagement and innovation co-creation, which is essential to produce knowledge capable of addressing the complexities inherent in symbiotic arrangements. Stakeholders range from business-to-business to the general public and will be engaged through Communities of Practice (CoP), immersive media experiences in multi-use play spaces, and through living labs which will contain specific location-based stories and visualizations driven by real data adding gamification elements (WP3). The project addresses the socio-political, regulatory and governance (pre)conditions for WSIS systems and services and will explore the conditions that enable or constrain the exchange of recovered water, energy and nutrients between industry sectors in the CE (WP4). All this work gears towards the development and demonstration of new WSIS business models and market opportunities, also supporting the emerging WSIS business ecosystem through a marketplace for Water, Energy, Materials in a WSIS. This marketplace is to be firmly embedded into Water Europe's real face-to-face water marketplace (WP5). Last but not least, ULTIMATE connects, creates synergies, and supports learning (WP6) in partnership with Water Europe and experts in science communication.

A dedicated set of transversal **WSIS Support Tools** will also be developed (WP2) that will support decision-making through presentation of the appropriate tools for individual solution assessment and system-wide evaluation stress testing. These tools will be embedded in the ULTIMATE website. The website will also provide user-friendly access to the TEB with information of e.g. outputs of the models, databases, KPIs, relationships based on user experience, etc.

Regarding the **ULTIMATE Marketplace** (task 5.5), several services will be developed such as linking to the TEB (WP1) and allowing users to explore the innovative technological solutions and showcases, finding collaborators and clients by building on

a 'match-making' functionality, providing personalised information to interested parties, funding innovation and supporting spinoffs.

To improve the visibility and commercialization of technologies and circular economy solutions and systems around resource use in the water sector, a general strategic approach is developed that includes target group assessment and definition of specific sets of knowledge products for case sites and technology application (WP2).

ULTIMATE Marketplace intends to share experiences and outcomes with stakeholder groups through its online platform that will be linked to existing networks, including Water Market Europe, the ICT4WATER cluster and WaterShare.

2.3. Types and formats of data generated/collected

Within WP1, data regarding the performance of the technical solutions proposed in the project will be collected as in-situ data by the site leaders. Additionally, a short description of the demo sites and some KPIs in terms of water, energy and materials will be requested by the cross-cutting technology groups (CTGs) for setting the baseline and quantifying the impact of the proposed system. A general description of each site will be collected through a template, where a description of the current and ULTIMATE systems, in terms of technologies used, volumes treated and outputs generated, as well as block diagrams will be included. Data for KPIs will be requested from the site leaders in a powerpoint presentation and discussed personally as well as via email and included in Milestone 5 (delivery in Month 6), which will be updated during the project execution.

WP2 will assess economic and environmental performance of individual WSIS technologies and associated risks, optimise the systems and make assessment and modelling results and tools available in a sustainable manner. In order to perform this, the following data will be collected:

1. technology specific data (e.g. system configurations, material and energy balances, quality of recovered products, potential microbial and chemical contamination, investment and operational cost elements, etc.) for baseline (i.e. status quo) and ULTIMATE configuration
2. infrastructure data (storage capacity, pipes sizes, etc.) for modelling the components of the physical and engineered system
3. historical data (water and material flows, energy, etc.). Historical data will be retrieved and converted into an interoperable and standard data format (such as OGC standards including WaterML 2.0, InfraGML, SAREF4WATER, etc.) and stored for further use in the WSIS support tools.
4. definition of the management boundary conditions: criteria and objectives, current and future scenarios. Management boundary conditions will be defined with water authorities/utilities and stakeholders within the CoP activities at the CS sites

5. data regarding the stakeholders needs and interests for the ULTIMATE toolbox will be collected as part of the CoP meeting. The main interest on the tools, the use cases for using it as well as the expected view of the results will be reported in an anonymized way. Once the tools are developed, they will allow the creation of user profiles giving the possibility to use the tool in a personalized way. Finally, the users will be able to provide their feedback on the returned results of the tool, aiming to collect this data and further improve the returned results.

The following data will be generated:

- Assessment of environmental impacts in various categories (e.g. global warming potential, primary energy demand, human toxicity, ⋯) of the applied technologies using defined LCA indicators.
- Semantically linked information forming a database to model and represent the interaction between different industrial processes and systems.
- Risk assessment will generate site specific uncertainty estimations to assess the probability of achieving the predefined objectives, e.g. reduction of microbial counts, compliance with of environmental legislation using recycling products. The data generated will be in the form of probability distributions as a function of system parameters (e.g. treatment time, yield)
- Optimised industrial ecosystem pathways, based on dilution-recovery cost relations for individual industrial processes (informed by T2.2.1) and the ontology mapping (T2.1), at the local and regional scale
- Model configurations combining inputs (standard or shocks including CC), systems (e.g. upstream changes, failure of components) with corresponding performance (e.g. circularity indicators of the water utilities, ecoefficiency, reliability, contamination risk, materials produced, profits, ecosystem services etc.) for stress testing the WSIS pathways (T2.3)
- KPIs that span the technological, environmental and socio-economic domains and are able to capture the performance of circular eco-innovations and trace its trajectory in the short, medium and long run.

Within WP3, data regarding stakeholders' views of the demonstrated circular water solutions will be collected as minutes of CoP meetings. The design, development and testing of the ULTIMATE experiences in WP3 will not require the gathering and processing of any personal data. The perspectives of the stakeholders will be reported in the CoP minutes, e.g. as anecdotal stories. In these minutes, all personally identifiable data will be anonymised. For the purpose of cross-fertilisation between the demo cases, the CoP reports will be stored and made available to the project partners in the ULTIMATE SharePoint. Furthermore, as part of the assessment of the immersive media experiences, data on people and their behaviour will be collected. This data typically consists of:

- Consent forms and demographic information
- People survey, interviews, well-being measures
- QoE, QoS data points

- Behavioural observations
- Affective or stress states using physiological measures such as ECG, GSR

This data will be handled and stored according to the GDPR (see section 6.3). Furthermore, ethical issues apply to collection, storage and use of these data. The ethical aspects and how these are handled in ULTIMATE are discussed in Deliverable 7.5 – Ethical Principles and Guidelines for Responsible Research and Innovation. For the purpose of cross-fertilisation between the demo cases, the CoP reports will be stored and made available to the project partners in the ULTIMATE SharePoint. Please refer to D7.1 for more details on the Sharepoint structure and use policies.

WP4 is expected to collect and process information regarding the socio-political and governance context for water, materials and energy in Water Smart Industrial Symbiosis systems and solutions. Relevant information, in the form of legal documents, reports, policy papers etc., will be collected and, where appropriate, will be included in the ULTIMATE project's knowledge environment (SharePoint). Furthermore, WP4 will collect information on the perception of the general public through surveys and interviews. These data will be handled and stored according to GDPR (see Section 6.3).

One main component of WP5 is the Marketplace, aiming at providing information on showcases collected in WP1, storing information of new initiatives and spin-offs in the field of innovation funding and, most important, collecting information referring to end users, technology providers and products. The Marketplace will build on and extend the Digital Marketplace developed in H2020 project NextGen, developed by KWR and NTUA, hosted and managed by WE. Selected information may be exported in other forms such as pdf (factsheets) or CSV (tables).

WP6 will collect information, as above, for the project's communication and dissemination purposes.

2.4. Reuse of Existing Data

The technology evidence base (TEB) developed within WP1 will build upon the technology evidence base developed in the NextGen project. The ULTIMATE TEB will be integrated in the NextGen TEB and will contain information about the circular economy solutions developed and demonstrated in ULTIMATE. Although it does not deal with industrial applications, water technologies from NextGen can be relevant within ULTIMATE as well, therefore, where of added value, NextGen data will be taken up. The data types included in the factsheets in the TEB are shown in Table 2.1, and an example of the structure for navigation in the TEB is provides in figure 2.2.

Table 2.1: Content of technology evidence base (TEB).

| Type of information	Data Format	
Name of technology	• Text	
Background	• Text • Image	• Text with short and long description (xx words) • Picture of products
Flow Chart	• Image	• Basic flow chart to understand the technology
Tags	• Text	• Type of product recovered, point of application
KPIs + technical parameters	• Text	• Data for performance and other aspects (y parameters with ranges)
Requirements	• Text	• List of parameters or boundary conditions for a favourable operation (parameters with ranges, qualitative information)
Legal and Regulatory information	• Text • Hyperlinks	
Business opportunities	• Text • Images	• Information on new business models
Case Studies	• Hyperlinks • Documents • Text	• Links to NextGen sites as examples, potentially including assessment results

TEB – Nutrients/Material
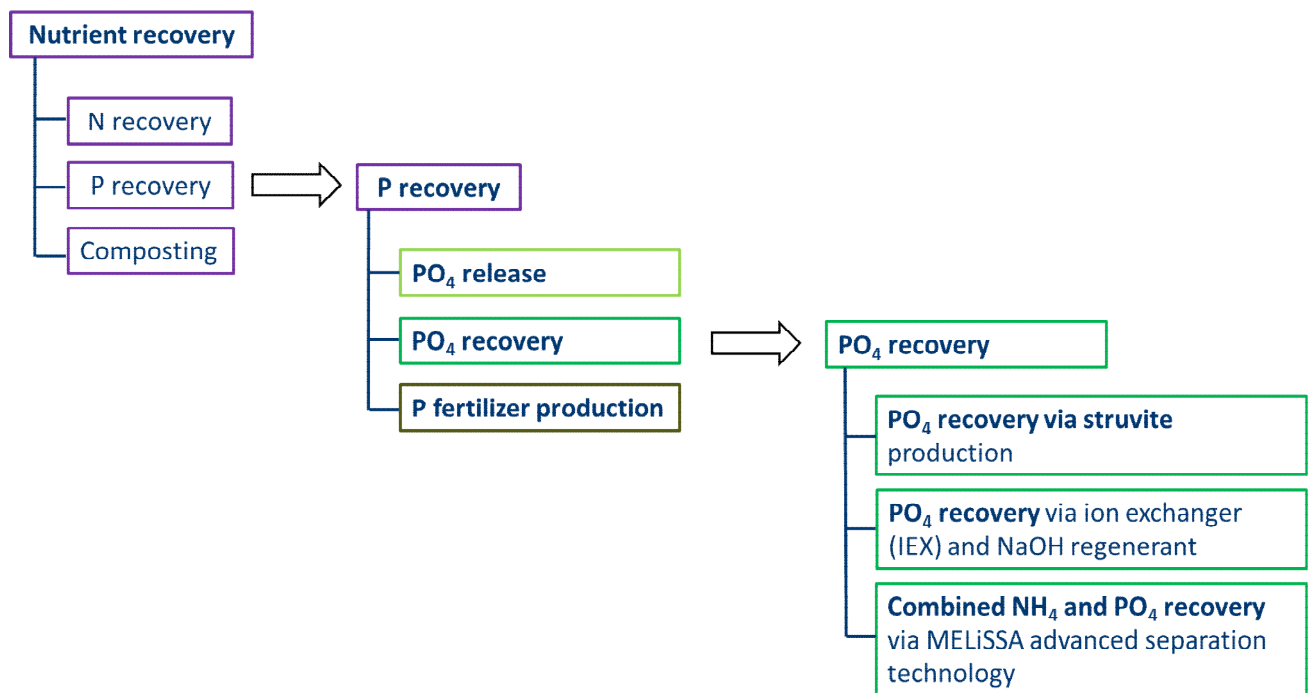
Figure 2.2.Example of structure in the NextGen TEB (still under development; status of October 2020).

Within WP2, a knowledge graph will be built to store information about the industrial processes. This knowledge-graph will be based on an adaptation of SAREF4WATR to include industrial domain, waste paths and also, materials paths. Moreover, it will also include the assessment and modelling work in Tasks 2.2, 2.3 and 2.4, data about the technologies of ULTIMATE will be used and returned to the users. Furthermore, the users will potentially have access to the data of the Technology Evidence Base that will be created as part of WP1.

The Marketplace (WP5) will build on the technology developed in the NextGen project. Special attention will be given in the collaboration with NextGen and as well as the collaboration with the projects B-WaterSmart, Water-Mining, REWAISE and WIDER UPTAKE which were funded under the same CE call as ULTIMATE, to ensure synergies and inclusion of results, if possible, in an integrated Marketplace.

2.5. Origin of the data

Data collected within WP1 will be obtained from in-situ monitoring of the performance of the technical solutions tested within ULTIMATE.

Data collected within WP2 will be obtained from in-situ monitoring of the performance of the technical solutions tested within ULTIMATE and from interviews with the contact persons of the sites and from the CoP of the sites as the users of the ULTIMATE WSIS tools. These data will function as input for the modelling and optimisation work performed within WP2, which will generate further data related to recommendations, predictions and also KPI information over the collected data to mention a few.

Data collected within WP3 will be obtained from stakeholders participating in CoP meetings, and the general public voluntarily participating in outreach events, serious games and augmented reality activities at the demo cases. We will ensure fully informed understanding of the implications of participation through information sheets.

In WP5 data will be collected from project partners in the first phase of the project. Additional information may be collected or accessed online by connecting the Marketplace with other established European platforms. In the final phase the Marketplace will be opened to registered users which may add further content.

2.6. Expected size of the data (if known)

In WP1 substantial datasets with system status information and measurements results are expected to be generated in all of the case studies. Furthermore, results from pilot scale and laboratory scale experiments will be created. The size of each of these datasets is currently not known.

In WP2 there will be data of about 10-15 participants at the CoP for the ULTIMATE WSIS tools. Furthermore, few hundreds of users are expected to use the WSIS tools and provide data. Substantial datasets with outcomes from the modelling, optimisation and re-design studies are expected. The size of these datasets is currently not known.

In WP3, the size of data collected is difficult to estimate but anyway substantial, as it relates to the input from approximately 4 to 6 stakeholder organisations participation in each of the nine CoPs (one for each Case Study), and thousands of citizens participating in the living labs events and three serious games / augmented reality activities.

In WP5 it is expected that the Marketplace database will store active users, their individual preferences (see also 3.2.1), technology providers and technologies in the order of hundreds.

2.7. Data utility: to whom will it be useful

Data collected within WP1 will be useful for the work of other WPs of the project, such as WP2 and WP5. WP1 comparisons of the different solutions tested will be used for engineering, water utilities and other specialized audience for selecting the best circular technologies to be applied in future projects.

The data collected and generated in WP2 will be used to assess economic and environmental performance of individual WSIS technologies and associated risks, optimise the systems and make assessment and modelling results and tools available in a sustainable manner.

Data collected within WP3 will be useful for the partners in the demo cases. The outcomes of the stakeholder collaboration in the CoPs and the public engagement in the Living Labs will benefit the devise, implementation and acceptability of the CE water technologies. Data collected within WP3 will be useful for WP4 of the project as well. Generic lessons from WP3 will be included in the EU roadmap for water, resources and energy in the CE.

3. FAIR Data

In general terms, the research data should be 'FAIR' that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard or implementation-solution.

3.1. Making data findable, including provisions for metadata

3.1.1. Discoverability of data (metadata provision)

All ULTIMATE subsystems, including the TEB, the Marketplace and the WSIS tools developed within the framework of the project, will be integrated into a common online system. They all give (open or limited) access to certain data categories. Thus, one way to increase the discoverability of the data is to develop HTML pages rich in <meta> tags which provide metadata about the HTML document and are parsable by other machines. Additionally, all subsystems, and mostly the integrated system of the three elements, will be linked to the project website. Efforts will be taken to establish also cross-links to other relevant websites.

ULTIMATE will encourage project partners to store datasets produced in open research repositories and obtain Digital Object Identifiers (DOIs).

3.1.2. Identifiability of data and use of persistent and unique identifiers

Objects stored in the ULTIMATE TEB will be associated with a unique and persistent ID, at the appropriate resolution (i.e. scale or aggregation level). The detail page of these objects can be accessed through a specific URL.

At the end of the project a snapshot of the whole database will be exported and stored in the research data repository Zenodo obtaining a DOI. Other categories of data produced by this project such as publications and presentations will also be uploaded to Zenodo or another research repository providing DOIs.

3.1.3. Naming conventions used
The proposed naming convention for datasets in the ULTIMATE project will consist of three parts:
- **A prefix indicating the data category (e.g. metadata, backup data, template)**

- **A short name for the dataset**
- **An integer indicating the version of the dataset**

The above information will also be used as a unique code for the dataset. Additional metadata such as the data provider and the creation date will be associated with the code and uploaded to the repository.

3.1.4. Approach towards search keyword

The ULTIMATE TEB and Marketplace will provide a variety of ways for data retrieval, such as full text search with approximate string matching and innovative advanced search capabilities based on the attributes and relations of various data categories, including keywords and tags. Additionally, georeferenced data (e.g. case studies) will be projected on a map enabling a geographical approach for the identification of these data and the navigation to detail pages. Other platforms which this project will use for certain data categories (e.g. Zenodo, GitHub) provide similar search capabilities.

3.1.5. Approach for clear versioning

Various information systems will have to be developed during the lifetime of the ULTIMATE project. Many of them will require the collaborative development of software applications, websites and other computer programs either from scratch or at least significant parts of them. The management and version control of the source code may become a significant problem, especially in large software development projects involving several developers.

The ULTIMATE IT team will systematically use Git as the source code management system for software development. Git will allow the collaborate development of the software providing effective distributed version control. It is easy to use and yet powerful and efficient to handle even large projects.

A copy of each Git repository will be uploaded ("pushed") to a publicly available Git server such as GitHub, Gitlab or Bitbucket. They are both well-established online servers which support distributed source code development, management, and revision control. They enable world-wide collaboration between developers and also provide some additional facilities to work on documentation and to track issues. GitHub provides paid and free service plans. Free service plans can have any number of public, open-access repositories with unlimited collaborators. Private, non-public repositories require a paid service plan, and free plans for academic/research projects. Bitbucket allows private repositories to be shared with up to five collaborators or with an unlimited number of users in case of academic/research projects.

In general, all datasets produced by ULTIMATE will receive a unique code, part of which will specify the version of the dataset.

3.1.6. Standards for metadata creation (if any)

Presently there are numerous standardisation activities including risk assessment and protection of a high level of services, although sometimes optional, is considered integrated in the entire operational process.

Other related Standards include:
- **ISO 9001:2015 provides an integrated approach to quality management - putting quality at the heart of business, touching on business resilience.**
- **ISO/IEC 27001 is an internationally recognised best practice framework for an information security management system.**
- **ISO/TR 37150:2014 Smart community infrastructures - Review of existing activities relevant to metrics addresses community infrastructures such as energy, water, transportation, waste and information and communications technology (ICT), focuses on the technical aspects of existing activities which have been published, implemented or discussed. Economic, political or societal aspects are not analysed.**
- **ISO 19115, Geographic Information – Metadata**
- **ISO 19110 Geographic information – Methodology for feature cataloguing**
- **ISO 19139, Geographic Information – Metadata -Implementation Specification**
- **OGC Catalog Service**

Efforts will be made that all textual information collected in the ULTIMATE project and especially all new information produced and stored in the ULTIMATE databases will be UTF-8 encoded.

3.2. Making data openly accessible

3.2.1. Which data will be made openly available, which data will not be disclosed and the rationale behind this

It is expected that most of the data produced or collected in the framework of the ULTIMATE project will be openly accessible. However, for some reasons, related with licensing issues, 3rd party data policy, and the use of personal or sensitive data, accessing may be restricted. The Marketplace is such an example, where efforts are taken to provide more targeted and personalised information to the users, utilizing advanced methodologies such as machine learning, semantic modelling, agent based modelling and other artificial intelligence techniques. As part of this process, information of the individual user's preferences will be collected and stored, to which only the Marketplace application will have access.

3.2.2. How the data will be made available

In ULTIMATE, various online tools will be developed that will provide access to data. Examples of such tools are the TEB, the Marketplace, the HMS simulation and stress testing platform, the interactive gamified visualisation tool, and the KPI assessment

tool. As more complete list of online tools will be provided in later versions of the DMP. These tools will be made available through the project website (access to project data) or the ULTIMATE online Marketplace. Specifically, the project's online environment will form part of Water Europe's infrastructure and as such will be developed and maintained beyond the project providing access to collected as well as to new material.

Additionally, at the end of the project a copy/snapshot of all relevant data as well as other data categories (publications, presentations, etc.) will be made available to open data repositories such as Zenodo.

3.2.2.1. Open access to publications
Any publications from ULTIMATE must be available as open access (as far as practicable possible). Open access to publications can be ensured either by publishing in Gold open access journals or Green open access journals.

Gold open access means the article is available as open access by the scientific publisher. Some journals require an author-processing fee for publishing open access.

Green open access or self-archiving means that the published article or the final peer reviewed manuscript is archived by the researcher in an online repository (e.g., on InnovationPlace), in most cases after its publication. Most journals within the social sciences domains require authors to delay self-archiving to repositories to 12 months after the article first being published.

In the ULTIMATE project, author-publishing fees for gold open access journals can be reimbursed within the project period and budget. There is, however, a very good selection of relevant gold open access and green open access journals available that do not charge author processing fees. Scholarly publication can take a very long time, and final acceptance of all submitted manuscripts may not occur before the end of the ULTIMATE project. For these reasons, we will prioritize to submit our work to gold open access journals without author processing fees, or to green open access journals.

3.2.3. What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
The data will be made available online through the WWW service. Thus, in order to discover, visualize or download the data only a common browser will be required from the user. Some online applications may require the user to register first before accessing the data. The format of downloadable files will be (solely or additionally) in an open standard such as CSV and PDF. Tools to access these files will be freely

available in the internet. User guides will be elaborated and provided for online environments, as well as for any non-intuitive tools produced within the framework of the project.

3.2.4. Association of data with metadata, documentation and code depositing

Several websites will provide access to data such as the following:
- **the project website (Access to project data)**
- **the ULTIMATE Marketplace (Products & Services, Stakeholders etc.), which will also provide access to the ULTIMATE Technology Evidence Base.**

Additionally, at the end of the project a copy/snapshot of all relevant data as well as other data categories (publications, presentations, etc.) will be made available to open data repositories such as Zenodo.

3.2.5. Providing access in case there are any restrictions

In order to gain access to the online Marketplace, a user will have to register to the Marketplace first and provide data regarding the user's background and fields of interest. The procedure is then simple and straightforward, involving the verification of the user data via email. Also, structured questions regarding specific technologies can be formed and a guided navigation with specific technological solutions will be provided.

3.3. Making data interoperable

3.3.1. Interoperability of data (data and metadata vocabularies, standards or methodologies followed to facilitate interoperability)

Datasets produced by the ULTIMATE project will be interoperable allowing data exchange and reuse. All systems will be user friendly, well documented and unless otherwise specified openly accessible. The ULTIMATE project will follow established European metadata vocabularies, standards and methodologies.

3.3.2. Use of standard vocabulary for all data types present in data set, to allow inter-disciplinary interoperability, or use of mapping to more commonly used ontologies

We will elaborate an ontology to determine waste utilization potential of industries and subsequent processes and systems to support WSIS. This ontology will integrate cross-domain indicators (water, energy, environment, food, etc.) with socio-economic indicators to improve the understanding of waste utilization potential at process, industry or industrial cluster level. This alignment between costs and processes will be complemented with the linkage of this information with the water-nexus ontology and

social modelling (using FOAF or SNS ontologies among others). The ontology will be delivered in the form of an online tool that allows interested industries to register their own water-related waste streams and identify promising opportunities for reuse at an industrial cluster and/or regional scale.

3.4. Increase data re-use (through clarifying licenses)

3.4.1. Data licencing to permit the widest reuse possible
Project partners intend to provide most of the data openly accessible, either immediately or after an embargo period. In these cases, the preferred license types are CC-BY (Creative Commons, Attribution) and CC-SA (Creative Commons, Attribution + Share Alike) and BY-NC (Creative Commons, Attribution + Non-Commercial).

3.4.2. When will data be made available for re-use. If applicable, specify why and for what period a data embargo is needed
Some data are already available online through the project websites. Other datasets will be uploaded to Zenodo as soon as they are complete and will be made immediately available through this service. Finally, for some data an embargo period will be necessary in order for them to be published first. It is estimated that for these data the embargo period will not last longer than until the end of the project.

Data that contains sensitive information about the operational and business processes of consortium partners will be classified as confidential, and will not be made publicly available.

Beyond commercial exploitation via existing partners, ULTIMATE is adopting an ambitious joint exploitation strategy, based on the creation of start-ups. Specific data that would compromise the possibilities to commercially exploit the outcomes of the ULTIMATE project may be classed restricted or confidential until a) a decision has been made on the relevance of this data for the valorisation process b) it's protection or the protection of the service or solution derived from it, have been properly protected (e.g. through a patent application)

3.4.3. Use of data produced and/or used in the project by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why
Almost all data produced and used in this project are available by third parties or will be made available after the end of the project. A small part of them will still remain

closed either because of licensing issues or because they refer to personal information (e.g. user profiles).

3.4.4. Data Quality Assurance Processes

Data validation process will be finalized by the end of the project.

Some aspects that will be considered are the following.

- **All parameters measured by the monitoring stations must be checked against extreme values in order to identify possible defect of instruments or another malfunction of the system. The responsible user will be notified in case of an issue.**
- **Upon data entry into the Evidence Base or the Marketplace several validation routines must apply informing the user in case the record is incomplete or erroneous. Such validation checks include missing mandatory values and data type check.**

Finally, several constraints embedded in the database should ensure the integrity of the data

3.4.5. Length of time for which the data will remain re-usable
All open data will remain re-usable for at least five years after the end of the project and maintained by the project partners. As suggested earlier, the project's online environment will form part of Water Europe's infrastructure and as such will be developed and maintained well beyond the project.

4. Allocation of resources

4.1. Costs for making data FAIR.

Most of the project partners apply software tools, technologies and protocols which comply to the FAIR principles and/or facilitate FAIR use of data. The use of WWW for various applications of the project documents this approach. Additional effort will be required at least for the following tasks:

- **Implementing various searching techniques (est. costs: 3 PM)**
- **Data conversion (est. costs: 2 PM)**
- **Establishing interoperability with other projects, Water Market Europe in particular (est. costs: 6 PM)**
- **Managing data in open research data repositories (est. costs: 1 PM)**
- **Developing and implementing a data management plan (est. costs: 1 PM)**
- **Implementing special interfaces (API) for accessing data through the Internet (est. costs: 2 PM)**

The costs will be covered by the project partners responsible for the respective tasks.

4.2. Responsibilities for data management

- **KWR** is the responsible partner for the overall coordination of the data management.
- **KWB** is responsible for developing and coordinating the collection of relevant data in the frame of WP1 and populating the Technology Evidence Base. KWB is supported in this role by EUT and UNIVPM.
- **EUT** is responsible for management of the activities developing the ULTIMATE WSIS support tools in WP2, which also involves **UNEXE**, **NTUA** and **KWR**.
- **KWR, NTUA and WE** are responsible for developing the ULTIMATE online Marketplace in WP5.
- For hosting and publishing the TEB, **ESCI** is responsible during the project and **WE** is responsible after the project.

Additional data managers such as the case study leaders at the various project partner demo cases (WP1) or Communities of Practice (WP3) take care of the other data produced locally.

4.3. Costs and potential value of long-term preservation

A snapshot of all relevant project data will be stored in the research data repository Zenodo upon termination of the project. This dataset will receive a DOI and will be preserved for the foreseeable future. However, the dataset will continue to grow even after the end of the project as real time measurements will keep coming from the

various project sites. Project partners responsible for the stations and other local stakeholders involved in their management and the exploitation of the data have an interest to keep these stations in operation as long as possible.

5. Data security

5.1. Data recovery and secure storage and transfer of data

Developing and establishing a backup and recovery plan for all ULTIMATE data is a crucial part of data security. While a general-purpose repository like Zenodo will store and publish the significant releases of the data produced in ULTIMATE, a backup plan takes care of the daily backups of all data. A variety of backup software exists for all operating systems, enabling manual and automated backup of the data. State of the art backup systems offer efficient storage of data (full and incremental backup, differential, backwards deltas) which keep track with older versions and even allow the reconstruction of the data at any single point in time.

From the perspective of the data security, we distinguish between the following data categories:

- **Data having no overall importance for the project. These data are often temporary and/or serve as an intermediate step for the production of other data. They are usually stored in various files. For this category of data, it is in the responsibility of the local system administrators to establish reliable backup plans, if necessary.**
- **Data stored in the project's website, in the ULTIMATE TEB, the WSIS Support Tools or the Marketplace will be backed up centrally in a remote backup server. The backup server will be located physically at a different site than the other servers and will be connected with them through the Internet. Thus, it will not be affected by any malfunction or damage which may occur at the site of each server. A filesystem snapshot utility based on rsync (e.g. rsnapshot) will be used to make periodic snapshots of local servers and transfer the data securely via ssh protocol to the remote backup server.**
- **Data collected by the monitoring systems at the various case study sites. These data can grow into large volumes depending on the size of the monitoring network, the number of the observed parameters and the recording time step. In case of high frequent observations, more frequent backups or establishing a replication system will be required. Depending on each case, monitoring data may be and stored in local and/or remote backup servers.**
- **The ULTIMATE Git repository stored in a remote server will serve as a means to backup software code produced during the project. An established service provider will be selected (e.g. GitHub or Bitbucket) guarantying long term preservation of the data. As the code is expected to be open source, its management will be taken care by the community.**

At the end of the project, a copy/snapshot of all significant and unrestricted data will be stored in the project's repository at Zenodo or published in open access journals.

The infrastructure of Zenodo has been developed and is supported by CERN which guarantees long term data safety and availability.

6. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

A number of ethics issues have been identified for the execution of ULTIMATE, specifically humans, non-EU countries, data protection and compliance with the GDPR and others. A response to each of these issues is given below.

6.1. Humans

Ethics issues in ULTIMATE are a subject of this deliverable, other deliverables (D7.5, 7.10 & 7.11) as well as the relevant framework and will continue to be developed and improved throughout the project.

As humans are involved in ULTIMATE, ethical issues related to their involvement, notably regarding recruitment and informed consent procedures need to be sufficiently described.

In D7.5, it is explicitly described how personal data as part of workshops, focus groups, interviews, CoPs, living labs and other dissemination activities in ULTIMATE should be collected, used and managed. These reports also propose measures to protect particular individuals and stakeholders, anonymise their contribution and minimise the risk of their stigmatisation from improper use of their data. As a rule, it is recommended that to the extent possible, only anonymized information is stored as part of the collected data.

D7.5 also describes the foreseen ethics considerations, related to the collection and processing of the data required for research and development of the ULTIMATE project. Issues such as data collection, informed consent, confidentiality, the code of conduct and the balance between data protection and freedom of rights have been identified. Information is also provided on procedures to be implemented for data collection, storage, protection, retention and destruction in compliance to EU Legislation.

In ULTIMATE, demo cases are being performed in close cooperation with local end-users, knowledge providers, technology providers, and water management agencies, thereby building local CoP groups. In ULTIMATE, cooperation through associate partnerships has been established. Data will be collected from participants in workshops, living labs, interactive play spaces, and other opinion elicitation activities. Personal data may also be obtained from already existing data files owned by the

different companies, industries and public administrations involved to the project, provided that said entities already poses the necessary consents to reuse personal data for research purposes. Each of the data sources used will be already complying with the data protection legislation applicable to each of the countries of origin, and in particular all data files should already be registered at the relevant data protection agency. All original data files have to be particularly compliant with the right to access, modify, cancel and object to further treatment of personal data.

6.2. Non-EU countries

The objective of ULTIMATE is to establish a market breakthrough of WSIS technologies as robust, effective, sustainable, and cost-efficient answers to the water, energy and material challenges in industrial context worldwide. It practically explores replication potential through adoption of systems and services by demo cases within the project and opens new global markets.

As part of this work, two of the case studies will take place in non-EU countries: Case Study 6 (Israel) and Case Study 7 (United Kingdom). These regions were selected based on the severe water resources problems encountered and the long-term working relationships that already exist between local stakeholders and ULTIMATE partners, particularly recognizing specific framework conditions, such as socio-economic, cultural and environmental characteristics. Addressing the social and governance challenges to uptake of WSIS solutions and services (WP4) and promoting dissemination and capacity building activities (WP6), encourages the sharing of knowledge with stakeholders. Data will be collected from, and shared with, consortium partners involved in these case studies. Furthermore, key partner and work package leader of WP3, NTNU, is located in non-EU member state Norway.

The consortium confirms that the ethical standards and guidelines of Horizon2020[1] will be rigorously applied, regardless of the country in which the research is carried out.

The ULTIMATE partners that are not EU members are located in Norway, Israel and the United Kingdom. According to Commission decisions, personal data can flow to and from Norway and the UK[2] without any safeguards being necessary. The Commission has also recognized Israel as providing adequate protection. Therefore, the consortium does not need to take any further action for transferring data to Norway, UK and Israel beyond the application of Directive 95/46/EC.

---

[1] See the following webpage for an up-to-date overview of the H2020 ethical guidelines and principles: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/ethics_en.htm

[2] At the time of writing the situation concerning exchange of information between the EU and the UK, and agreements governing this exchange, are not clear beyond 31 December 2020. The DMP will be updated should there be a change in the situation.

6.3. Data protection and compliance with the GDPR

The ULTIMATE consortium will inform EASME regarding each activity what data will be collected stored and processed as a result of these interactions. In those cases where personal data is involved, detailed information is planned to be provided:

1. **on what personal data is collected, stored and processed**
2. **on the recruitment process, inclusion/exclusion criteria for participation**
3. **on privacy/confidentiality and the procedures that are implemented for data collection, storage, access, sharing policies, protection, retention and destruction during and after the project**
4. **on how informed consent is pursued**
5. **if application/s need to be filed with a local/institutional ethics review bodies (if personal data is being collected) and if yes, which bodies / where / when.**

Data will be collected in accordance with the General Data Protection Regulation (GDPR), a Regulation describing precisely the rights of data subjects and the obligations of data controllers and processors. The GDPR affects mainly the interaction with citizens in WP3 (interactive place spaces, living labs) and WP5 (Digital Marketplace) as there are expected to be the main components where information related to personal data is collected.

6.3.1. Work Package 3 – play-spaces and living labs

The design, development and testing of the ULTIMATE experiences in WP3 will not require the gathering and processing of any personal data. The ULTIMATE WP3 is committed to ensure that all the activities carried out will comply with ethical principles and relevant national, EU and international legislation, such as the Charter of Fundamental Rights of the European Union and the European Convention on Human Rights. For data protection and privacy, we will adopt a data minimization policy, collecting and processing only the data that is strictly necessary for running the activity. Participants will be informed about how and by whom data will be processed and stored and about their rights to access, modify and erase their personal data and about the procedures to enforce this. All data collected and processed will be in compliance with privacy. They will be hosted at protected servers at NTNU and thus will be submitted to the compliance with the NSD, as well as to the General Data Protection Regulation. Data protection measures will be applied with regard to data collection, storage, retention, destruction, privacy and confidentiality. Templates of the informed consent forms and information sheets with regard to processing and protection of personal data (in language and terms intelligible to the participants) will be kept on file on NTNU servers. All ethical approvals/opinions/notifications will be handled by the NTNU DPO, mr. Thomas Helgesen (M), Director Organization.

6.3.2. Work Package 5 – Digital Marketplace

Persons registering for getting access to the Marketplace are providing their consent through the online form using clear and plain language. Users will be able to withdraw their consent at any time, get informed of the data the system has collected and stores about them and apply to export them or permanently erase them from the system. Sensitive data collected by the Marketplace will be encrypted before storing them in the database.

6.3.2. Work Package 5 – Digital Marketplace

7. Other

Refer to other national / funder / sectorial / departmental procedures for data management that you are using (if any)

Up to now there is no other procedure for data management in the framework of the ULTIMATE project, apart from the ones that have been already described in the present document.

8. Conclusions

Most of the data produced or collected in the framework of the ULTIMATE project are generally expected to be openly accessible. Exceptions may apply for reasons related with licensing issues, third party data policy, and the use of personal or sensitive data, thus accessing this information may be restricted.

In particular, all ULTIMATE subsystems, including the Technology Evidence Base, the WSIS support tools and the Marketplace, will be integrated into a common online system, which will be giving open or limited access to certain data categories.

Additionally, datasets produced by the ULTIMATE project will be interoperable allowing data exchange and reuse. All systems will be user friendly, well documented and unless otherwise specified openly accessible. The ULTIMATE project will follow established European metadata vocabularies, standards and methodologies. All open data will remain re-usable for at least five years after the end of the project and maintained by the project partners.

A backup and recovery plan will be developed and established for all ULTIMATE data as it comprises a crucial part of data security. A number of ethics issues have been identified for the execution of ULTIMATE, specifically humans, non-EU countries, data protection and compliance with the GDPR and others.

In conclusion, the benefits of a well-designed DMP not only concern the way data are treated but also the successful outcome of the project itself. A properly planned DMP guides the researchers first to think what to do with the data and then how to collect, store and process them. The process of planning is practically a process of communication, increasingly important in a multi-partner research. The characteristics of collaboration should be accordingly harmonised among project partners from different organisations or different countries.