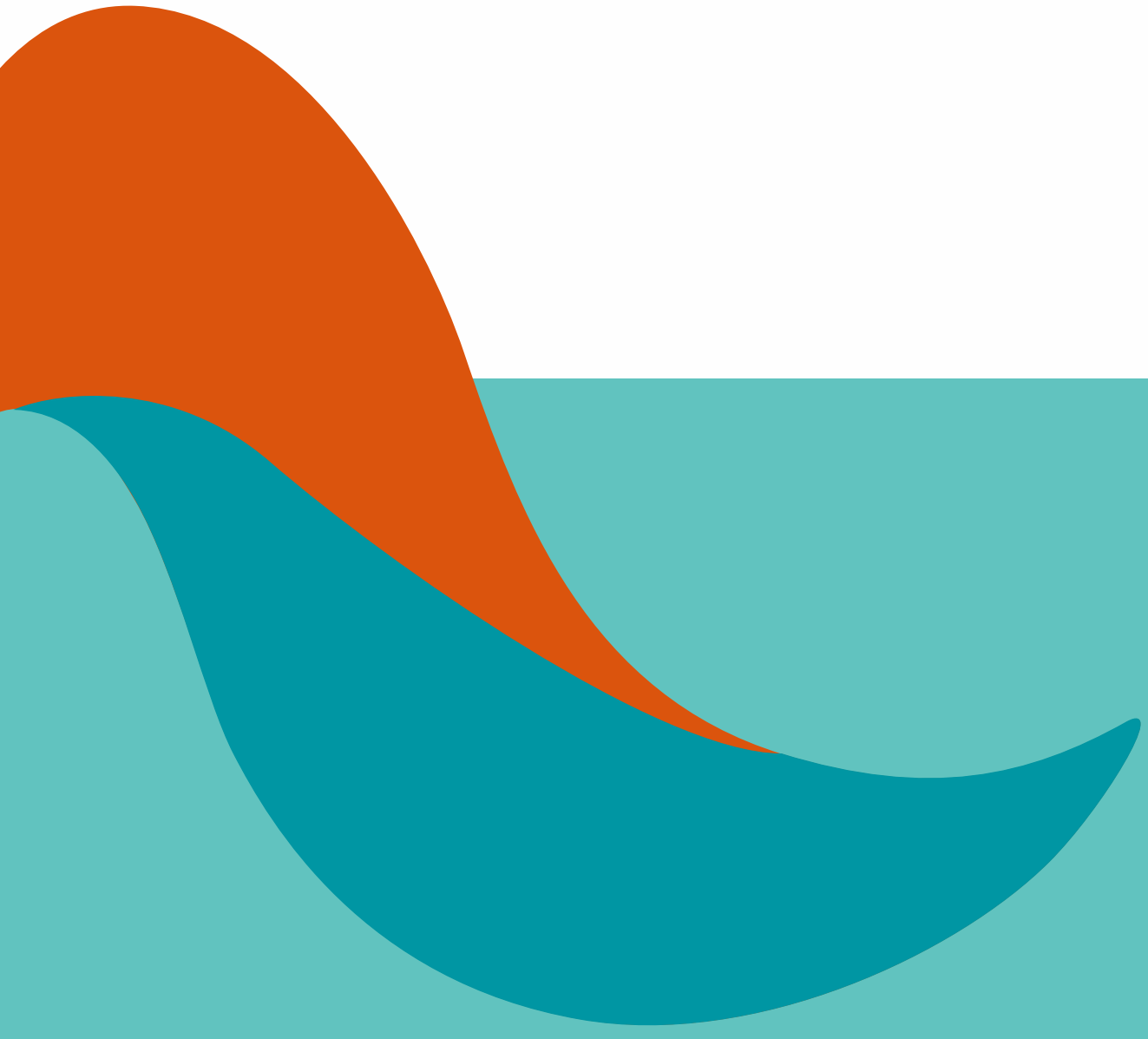
D7.2 Initial Data Management Plan

Eloy Hernandez (EUT)

June 28th, 2024



Introductory Table

Project Ref. No.	HORIZON-CL6-2023-CLIMATE-01-2; GA No. 101136598
Project Title	Reliability and effectiveness of integrated alternative water resources management for regional climate change adaptation
Duration of the Project	2024-01-01 to 2027-12-31 (48 months)
WP/Task:	WP7 / T7.4
Document due Date:	30-06-2024
Actual Date of Delivery	28-06-2024
Leader of this Deliverable	EURECAT
Dissemination Level	PUBLIC
Document Status	Submitted

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101136598. This document reflects only the views of RECREATE consortium, neither the European Commission nor any associated parties are responsible for any use that may be made of the information it contains.

Deliverable Information Sheet

Version	Date	Author	Document history/approvals
0.1	14-06-2024	Eloy Hernandez (EUT)	Draft version. First version of the document, prior to internal RECREATE members review
0.2	26-06-2024	Lydia Vamvakeridou-Lyroudia (KWR)	First Review
0.3	26-06-2024	Dionysios Nikolopoulos (NTUA)	Second Review
1.0	27-06-2024	Eloy Hernandez (EUT)	Final version for submission
2.0	28-06-2024	Digu Aruchamy (EUT)	Final Quality check before submission

Executive Summary

This document presents the initial version of the Data Management Plan (DMP) on open access data handling (see Box 1) defined for RECREATE. The aim of the document is to consider the many aspects of data management, data and metadata generation, data preservation- mainte-nance- 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 Scien-tific Publications and Open Data Access to Research Data in H2020. Complementary to this document, we have decided to use the Argos Tool (see Box 2) to maintain the DMP online and dynamically for the entire duration of RECREATE. Specifically, the corresponding online version of the datasets utilised in RECREATE will be available in the following URL corresponding to the public online version of the DMP (see Box 2)

Box 1. Open Access

Open access (OA) refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable. 'Scientific' refers to all academic disciplines. In the context of research & innovation, 'scientific information' can mean: (1) peer-reviewed scientific research articles (published in scholarly journals) or (2) research data (data underlying publications, curated and raw data).

Box 2. RECREATE DMP IN ARGOS TOOL

ARGOS Tool is an online tool to create, link and share data management plans. It is developed by OpenAire and enables the automation of the process of cataloguing and sharing data between researchers, communities, and funders. Moreover, ARGOS TOOL also permits shareing data according to common standards and at the end, make the DMP machine-actionable . For RECREATE, we have elabo-rated a data management in the following URL:
https://argos.openaire.eu/plans/overview/b367768d-5c88-4af8-9554-c204fce76dc7

The sections below present the lifecycle, responsibilities, review processes and management policies of research data, produced during the execution of RECREATE. The DMP reflects the agreement of the RECREATE consortium as well as the adopted measures concerning the control, protection, distribution, and maintenance of the produced data.

For RECREATE, the DMP is defined as “the data management strategy enabling data sharing, data privacy and data ownership. Description of how data will be handled during the project, and even after the project is completed, describing what data will be collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved”. Since the beginning of the project, the following processes and procedures for data management procedures are established:

- Data governance, such as standards management and guidelines.
- Data architecture, analysis, and design including data modelling.
- Data maintenance, administration, and data mapping across building blocks and solution 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 trans-fer, master data management, reference data.
- Document, record, and content management.
- Metadata management, i.e., metadata definition, discovery, publishing, metrics, and standardization.

Table of Contents

List of Abbreviations

AWR	*Alternative Water Resources*
CA	*Consortium Agreement*
CoP	*Community of Practice*
CS	*Case Study*
DMP	*Data Management Plan*
DoA	*Description of the Action*
EOSC	*Open Science Cloud*
EUT	*Eurecat Technology Centre*
FAIR	*Findable, accessible, interoperable, re-usable*
GA	*Grant Agreement*
IoT	*Internet of Things*
IPR	*Intellectual Property Rights*
OA	*Open Access*
OAI-PMH	*Open Archives Initiative Protocol for Metadata Harvesting*
ODC	*Open Data Commons*
PDDL	*Public Domain Dedication and Language*
SAREF	*Smart Applications REFerence*
WP	*Work Package*

List of Figures

List of Tables

1. Introduction

1.1 Scope

This document is devoted to the description and publication of the initial version of the Data Management Plan, a work performed inside WP7- "Project Management", and specifically, in Task 7.4 entitled "Data Management". The present document corresponds to the series of deliverables derived from D7.2.- "Initial Data Management Plan" in which the initial version is published in M6. This will be a living document and will be continuously updated during the project period. The updated DMP versions will be submitted as public deliverables in M24 (D7.3 Mid-project data management plan) and in M48 (D7.5 Final data management plan).

Box 3. Data Management Plan

A Data Management Plan (DMP) is a key element of good data management; it describes the data management life cycle for the data to be collected, processed, and generated by a Horizon Europe project.
As part of making research data findable, accessible, interoperable, and re-usable (FAIR), a DMP should include information on: (i) the handling of research data during and after the end of the project, (ii) what data will be collected, processed, and generated, (iii) which methodology and standards will be applied, (iv) whether data will be shared/made open access, and (v) how data will be curated and preserved (including after the end of the project). A DMP is required for all projects participating in the extended ORD (Open Research Data) pilot unless they opt out of the ORD pilot; however, projects that opt are encouraged to submit a DMP on voluntary basis.

The DMP provides a description about the procedures on how the research data is collected, processed, and generated. The DMP (see Box 3) establish the procedures on how to handle these data along the project and lastly, after finalization. Hence, the document (and the subsequent versions of it), represents the consortium agreement on the data plan and also a description of the main standards and methodologies that has been established for data collection, generation, sharing and preservation.

This document follows the template provided by the European Commission on DMP structure and guidelines[1]. This documented DMP is complemented also, by the publication of the datasets under the Argos Tool (see Box 2) published by Open Aire.

Box 2. RECREATE DMP IN ARGOS TOOL

ARGOS Tool is an online tool to create, link and share data management plans. It is developed by OpenAire and enables the automation of the process of cataloguing and sharing data between researchers, communities, and funders. Moreover, ARGOS TOOL also permits shareing data according to common standards and at the end, make the DMP machine-actionable . For RECREATE, we have elabo-rated a data management in the following URL: https://argos.openaire.eu/plans/overview/b367768d-5c88-4af8-9554-c204fce76dc7

The interrelation of the DMP with Open Aire allows RECREATE to be part also in the Open Research Data Pilot[2] contributing as well to the Open Science Cloud (EOSC)[3]. All of these aspects have been materialized in this initial version of the DMP since the M01 where following action will be done:

- Regular update of the DMP (online and the present document) will be provided at **M6 (D7.2)** with iterations on **M24 (D7.3), M48 (D7.4).**
- The online DPM will add datasets periodically, as a Dataset is ready to be published.

1.2 Structure of the document

The DMP document has been structured according to the following sections:

- **Initial section** Is mainly devoted for the executive summary of the document.
- **Section 1** Is the introductory chapter, which provides the scope of the deliverable and the main outline of the document.
- **Section 2** Contains information about digital datasets generated or collected in RECREATE for each of the WPs and, the sections devoted to the preservation data mechanisms established within the project.
- **Section 3** Contains information about the FAIR data for RECREATE and will be updated at same time as the project evolves.

[1] Guidelines on Data Management in Horizon Europe, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[2] Open Access to Scientific Publications and Research Data in Horizon Europe Guidelines, https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

[3] European Open Science Could, https://eosc-portal.eu/

- **Section 4** Focuses on the allocation of resources to maintain the datasets and digital assets elaborated within RECREATE.
- **Section 5** Section devoted to the data security aspects.
- **Section 6** Address issues related to ethical aspects.
- **Section 7** Contains other related issues related to the data management.
- **Section 8** Contains the conslusions and the future work within the DMP of RECREATE from M06 to the end of thr project.

2. Data summary

2.1 Purpose of the data collection/generation and relation to the objectives of the project

RECREATE envisions to improve water management in water-scarce regions. The primary aim is to enhance the resilience of water supplies and protect natural water resources by facilitating the assessment and inclusion of Alternative Water Resources (AWR) in water management planning. This involves collecting and generating data on water availability, quality, and demand, which are essential for evaluating the current state and potential improvements in water resilience. By gathering data on the environmental impact, cost-efficiency, and health implications of various water management practices and technologies, the project supports the development of sustainable strategies for protecting natural resources.

RECREATE seeks to increase awareness and acceptance of AWR by creating an open access repository of knowledge and data, thereby promoting their fundamental role in climate change adaptation. This repository will be encoded into a Digital Decision Support Framework (RECREATE_WT), which integrates comprehensive assessments of water availability, demand, and the performance of innovative treatment technologies. The data collected will enable RECREATE_WT to provide a decision matrix for co-creating adaptive water management pathways tailored to different stress scenarios and local contexts.

Furthermore, RECREATE will validate and demonstrate its digital solutions through four regional Case Studies (**Error! Reference source not found.**- Actions related to the Case Studies in WP5). These c ase studies will focus on the digitalization of the water cycle for integrated water management, showcasing the cost-effectiveness, reliability, and safety of using new AWR. By utilizing full-scale or already available pilots from previous or ongoing projects, RECREATE aims to optimize efforts and maximize the project impact.

Table 1 Pilot sites and CS leading partners and data sources

CASE STUDY	ASSIGNED INSTITUTION	DATA
North Holland (The Netherlands)	PWN	• Physical data/availability ○ Industrial abstractions and discharges from/to groundwater, and surface water and rainwater harvesting, and discharges on the sewage system. ○ Agricultural abstractions of surface water and groundwater

		○ Sewage water treatment plants (capacity) • System Dynamics Model ○ Evaluations of multiple scenarios for AWR management ○ Insights on how adaptations may propagate through the regional water system • Historical data ○ Climate scenarios ○ Socio-economic scenarios • Climate variables ○ Salinity levels ○ Precipitation patterns • Climate scenarios ○ Dutch national meteorological institute KNMI • Measurements ○ Water quality
Kalundborg (Denmark)	KCR	• Physical data/availability ○ Flowrates of main existing water and wastewater streams ○ Chemical composition of main wastewater streams (COD, TN, TP, Sulphur) ○ Energy and chemicals demand of existing WWTPs ○ Weather data: https://open-meteo.com/en/docs/historical-weather-api
Syros (Greece)	DEYAS	Water Consumption and Demographic Data • Consumption/Population Data per consumer group (Permanent residents, tourism, farmers) • Consumer Demographic Data (number of households, household members, income, age) • Energy and Water Bills as Time Series: (daily or monthly step) and tiered tariffs Climatic data

		• Historical time series of temperature and precipitation (Daily timestep)
		• Climate Data Projections: Downscaled IPCC scenarios in the form of time series of the climatic variables for the predetermined simulation horizon (i.e. 2024-2050,2050-2099)
		
		<u>Technical Data</u>
		• System Operation Rules (optional)
		• Desalination:
		  ○ Number and Processing Capacity (Average and Maximum Daily production)
		  ○ Time series of desalinated water production (Daily)
		
		• Wells:
		  ○ Number of installed units
		  ○ Maximum pumping capacity
		• Volume and Location of Reservoirs
		• Water Treatment Plants - Water Treatment Units:
		  ○ Location
		  ○ Maximum Capacity (Daily - Design)
		  ○ Storage Capacity
		• Wastewater Treatment Plants:
		  ○ Location
		  ○ Treatment Capacity (Maximum Daily - Design)
		  ○ Quality of produced water
		  ○ Time series of Treated Water Production
		• Water Distribution Network:
		  ○ Length (Approximate-Cumulative)
		  ○ Losses

Costa Brava (Spain)	ICRA	• Census Data • Tourism • Surface Water quantity and quality • Groundwater quantity and quality • Geological data • Abstraction wells • GIS Data (Agriculture, Urban, Infrastructure,...) • Land use, land cover, soil sealing • Water distribution network • Water treatment plants • Meteorological data (Rainfall, Runoff, Temperature) • Energy demand of WWTP and water reclamation plants

RECREATE consortium has identified the following datasets, inputs and their potential contributions to the consecution of such objectives from each case study:

Table 2 Datasets generated

#	INPUT DATASET	URL	CASE STUDY
1	Weather data	https://open-meteo.com/en/docs/historical-weather-api	Metereological data from Kalundborg
2	Helenic metereological services	http://emy.gr/emy/en/climatology/climatology_city	Metereological data from Syros
3	Meteo.cat	https://www.meteo.cat/	Temperature and precipitation data for Costa Brava
4	Idescat (Catalonian Statistical Institute)	https://www.idescat.cat/?lang=es	Provides data on available potable water, distributed water, losses, residual water and reused water. Relevant for Costa Brava
5	EDO - European Drought Observatory – Copernicus	https://drought.emergency.copernicus.eu	Earth observations. Relevant for all the CS

6	Daily observations of the Earth Ecosystem – Copernicus	https://www.copernicus.eu/en/access-datahttps://open-meteo.com/en/docs/historical-weather-api	Earth observations. Relevant for all the CS

(*) This table will be updated according to the project progress

2.2 Types and Formats of data generated/collected

Considering the Table 2, the datasets generated in RECREATE are related with:

- data sets referred to project publications (deliverables and papers)
- curated and/or raw data collected produced during the project.

Within RECREATE, datasets are primarily identified and collected through several Work Packages (WPs). These are the WPs and Tasks that are relevant for the DMP:

- **WP1 (Risks and Potentials of Alternative Water Use):** Assess climate change impacts, regulatory barriers, and health risks associated with alternative water resources.
- **WP2 (Stakeholder Engagement):** Engage stakeholders to foster community involvement and acceptance. Stakeholders requirements for the WP3 tool.
- **WP3 (RECREATE_WT):**
  - **Task 3.1:** Co-design the RECREATE_WT visualization tool to enhance user experience and decision-making through open source platforms like D3.js and HighCharts.
  - **Task 3.2:** Design a stable, cyber-secure, and interoperable architecture integrating FIWARE and extending data models for FAIR data sharing.
  - **Task 3.3:** Deploy microservices on the FIWARE platform for data preprocessing, storage in a time-series database, and visualization through an open-source dashboard.
  - **Task 3.4:** Streamline integration of RECREATE_WT outputs into JRC tools, ensuring compatibility with existing models like LISFLOOD and POLES for EU-wide smart water management studies.
- **WP4 (Strategic Planning):** Analyze policy options and frameworks for integrating alternative water resources into regional strategies.
- **WP5 (Case Studies):** Implement and evaluate alternative water resource solutions across diverse bioregions and water contexts.

The datasets collected through WP3, especially from tasks focusing on data integration and visualization, are crucial for supporting decision-making processes across multiple WPs. These datasets are processed and stored to facilitate scenario simulations and model outputs, ensuring interoperability and compliance with EU standards and JRC tools.

WP3 also uses information to establish a user-friendly visualization of the information according to the stakeholders. Considering this interconnection between WPs and informational flow. Under the conformation of these vase studies and deployment of the RECREATE_WT tool, there will be generated curated datasets, datasets coming from the application of AI driven tools.

Considering the raw data underpinning the published work, constitute the main research data sets that will be made publicly available if the Authors/Data providers give their permission. In cases where release of complete raw data sets is impossible due to, for exam-ple, privacy or industrial/personal

data concerns (e.g., industrial specific process information that could compromise the infrastructure), RECREATE consortium will stablish and maintain an exploitation plan of the data collected and generated in the following table:

Table 3 Details about the data generated and collected

#	IDENTIFIER	SIZE	TYPE	FORMAT

(*) This table will be updated according to the project progress

2.3 Specification about existing data reused

Although RECREATE aspires to produce its own curated datasets (outputs), it will also use in the most suitable way any existing data that can contribute to its purpose. In this regards, the following data will be used:

- Weather and climate data on different time horizons (short-medium, seasonal, and decadal projection) provided by cutting edge data climate services (e.g. Copernicus Climate Change Service (C3S))
- Climate data. Depending on CS, i.e CS Costa Brava use Meteocat as weather forecast data provider.
- Alternative water resources and water quality data from JRC tools.

In the scope of RECREATE project is not envisioned to re-use datasets from other consortium initiatives, if during the project this is required that information will be collected in the next table:

Table 4 Re-used datasets from other consortium initiatives

#	DATA ASSET	PREVIOUS INITIATIVE	CONTRIBUTION RECREATE

(*) This table will be updated according to the project progress

2.4 Origin of the data

Considering the main digital assets to be elaborated within RECREATE, the following table represents the origin of data needed to build and deploy the different tools. It is worth to mention that we also included reusable semantic models (like SAREF) and smart data models to ensure semantic interoperability between the digital assets.

Table 5 Data Origin of the different tools in RECREATE

#	TOOL	DATA ORIGIN
1	EPANET	• Hydraulic simulation • Water quality • Network analysis
2	IJsselmeer chloride prediction model	• Output of IJsselmeer chloride prediction model ('bakjesmodel')

(*) This table will be updated according to the project progress

2.5 Expected size of data

It is expected raw or text data size to be manageable in terms of storage capacity. Data like geospatial maps it is likely to require more storage capacity than other data.

As considered in the abovementioned sections, RECREATE will consider the publication and use of raw data, processed data and also output data coming from modelling. Based on these types of data detailed in Table 2 and Table 3, the potential aggregated data size is depicted in the follow-ing table:

Table 6 Overall data size by digital asset and dataset

#	DIGITAL ASSET TYPE	OVERALL SIZE
1	Historical Raw Data	(*) TBD in future versions
2	Dynamic Models & Simulation	(*) TBD in future versions
3	Real-Time Monitoring Platform	(*) TBD in future versions
4	Stakeholder engagement and Visualization	(*) TBD in future versions
5	Climatic change scenarios	(*) TBD in future versions

2.6 Outline the data: to whom will be useful?

The data to be generated within RECREATE is relevant for increase the knowledge to resilience assessment to climatic change for public and private use. Most of the generated data and digital assets will be for public and societal use and corresponds to data that will be publicly released in Zenodo to be used by scientists, industries, authorities and other interested parties in relation to the regionality. At this moment, the Zenodo has been opened and the next step will be to upload corresponding datasets prone to make them publicly available. Complementing this information, IT and digital tools development code will be available in different Github/Gitlab repositories. Moreover, final versions of these digital assets will be also published inside the RECREATE web-site. Considering these aspects, the following table reflect main stakeholders that could use the different data assets:

Table 7 Data utility by stakeholders

STAKEHOLDERS	DATA UTILITY
Scientist	Perform research and newer innovations in efficient use of resources and industrial symbiosis through the use of reliable datasets from different industries and regions.
Authorities	Knowledge sharing about zero waste and industrial and process symbiosis in different EU regions. These knowledge sharing will be driven from the reference digital architecture to be provided in RECREATE. Moreover, the information could be relevant from them to perform an overall vision about circular economy.
Environmentalist	Perform research, indicators and impacts in relation with circular economy through the use of reliable data sets from different representative regions in EU.
Industrial Stakeholders	Knowledge exchange about different industrial stakeholders in terms of resource efficiency, optimization and industrial symbiosis with the empowerment of circular economy.
Policy Makers	Establishment and reinforcement of their knowledge to develop adaptation measures, for improving water governance and management
Society at large	Increase awareness on the importance of circular economy in the fight against climate change. Also, it is relevant for this societal pillar the interaction with intuitive and user-friendly visualization environment to support future envisions on co-creation projects and expansion to other industrial companies.

3. FAIR data

This section should be understood as a living section that will be further updated in future iteration as the project evolves, and more input in terms of data and data reporting comes from the RECREATE partners.

Intellectual Property Rights (IPR) management in RECREATE is a substantial part of its data management plan. Usually, data content and their system are treated as one parameter, but when the matter comes to IPR, a distinction between the databases and data content is of outmost importance. It is imperative for other users to know how they can reuse both the data collected, assembled, or generated and the databases where these are included.

The Open Data Commons[4] group developed the following tools to govern the use of data sets. The three ODC licenses are:

- Public Domain Dedication and License (PDDL): This makes the use of the database and its content free to the public domain.
- Attribution License (ODC-By): Users can make use of the database and its content in new and different ways, but they need to provide an attribution to the source of the data and/or the database.
- Open Database License (ODC-ODbL): ODbL stipulates that any use of the database must provide attribution, and any new outcomes must use the same terms of licensing (also an unrestricted version of the new product must always be accessible).

Considering aspects related with the ownership of data combined with Open Research Data in RECREATE, the following license are prone to be applied to the digital assets to ensure maintenance of the provenance and background parties involved:

Table 8 License linked to the RECREATE data assets

DIGITAL ASSET	POTENTIAL LICENSE
LISFLOOD (JRC tool)	Free Software licence

[4] https://opendatacommons.org/

LUISA Territorial Modelling Platform (JRC tool)	Non-Free Software licence
Wotter Model	(*)To be Completed
hydroSHEDS	(*)To be Completed
Data-Driven Models for Groundwater Availability	(*)To be Completed
UWOT	(*)To be Completed
SUGGEREIX platform	(*)To be Completed
System dynamics model	(*)To be Completed
QMRA	(*)To be Completed
QCRA	(*)To be Completed
LCA	(*)To be Completed
Agent Based Models	(*)To be Completed
MODFLOW	(*)To be Completed
SWAT	(*)To be Completed
DWSIM	(*)To be Completed

(*) This table will be updated according to the project progress

3.1 Making data findable, including provisions for metadata

As depicted in the previous sections, RECREATE will publish open data coming from the scientific results and outcomes in Zenodo. Moreover, RECREATE will also publish the datasets in the Argos Tool (digital version of the DMP) initially created during this initial phase of the project. Taking advantage of these initiatives, the RECREATE consortium could ensure the research data findable in agreement with the Horizon Europe Open Access Mandate.

Hence, all uploads in Zenodo will be enriched with standard Zenodo metadata, including the following information:

- Publication Date
- DOI
- Grants agreement number
- Project acronym
- License
- Versions
- Publisher

By means of publishing information in Zenodo and inside the Argos Tool, RECREATE also ensure the information is indexed in Open Aire and then contribute to the open research at European level. This section will be updated during next iterations to provide detailed information on how data will be made discoverable, and more specifically:

- Discoverability of data (metadata provision)
- Identifiability of data and refer to standard identification mechanism
- Use of persistent and unique identifiers such as Digital Object Identifiers
- Naming and conventions used
- Approach towards search keyword
- Approach for clear versioning
- Specify standards for metadata creation
- Type of metadata created and how

3.1.1 Making data openly accessible

All data, information, and knowledge considered relevant for the scientific community will be open access to peer reviewed scientific publications resulting from the project through Green OA and

Gold OA, following Horizon Europe recommendations. In this regard, data will be shared in relation to (i) publications (deliverables and papers) and (ii) curated and/or raw data. For the data linked to

scientific publications, the publication will serve as the main piece of metadata documentation for the shared data. When this is not adequate for the comprehension of the raw data, a report will be shared along with the data explaining their meaning and methods of acquisition. However, for both data categories the metadata standard structure of the data repository (tentative an FTP server) will be used.

3.1.2 Making data identifiable

All the available datasets will be uploaded in Zenodo, Argos Tool and EOSC[5]. In this regard, dataset's reference and naming will be implemented to employ a standard identification mechanism for each data set according to the metadata standard implemented. Zenodo (a popular repository for research data, will be extensively exploited throughout the project) assigns all publicly available uploads a Digital Object Identifier (DOI) to make the upload easily and uniquely citable. Zenodo supports harvesting of all content via the OAI-PMH protocol.

Similar aspects are present inside the Argos Tool (digital version of the DMP). Inside Argus Tool, the general aspects of the DMP are linked with specific contributions using ORCID. The general information of the DMP is complemented with information about the datasets prone to be published. The information related to these datasets corresponds to the main FAIR principles. The corresponding datasets published are simultaneously shared with Zenodo and Open Research Data Cloud. Therefore, the main metadata adhered to the datasets are the same as exposed for the Zenodo case.

[5] https://open-science-cloud.ec.europa.eu/

3.1.3 Naming and conventions used

At this stage of the DMP, there is no defined naming conventions for the files due it will depend on the data available in the pilot cases. Nevertheless, the following general naming convention will be discussed in the following months and next versions of DMP document:

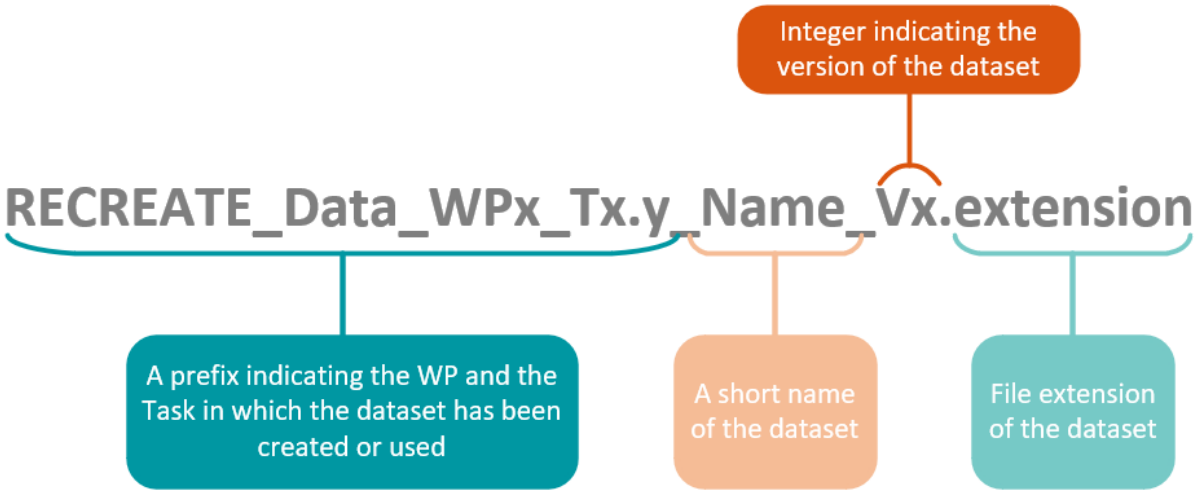


Figure 1 Tentative structure to name the files inside the repository

3.1.4 Approach towards search keywords

This section will be updated on next iterations of this Deliverable to provide detailed information on search keywords depending if data will be collected to be re-used. This section refers to whether the data will be accompanied by a set of keywords that make it easier to find, understand, and reuse the data:

- **Search Keywords:** These are specific terms or phrases associated with the data that help in categorizing and indexing the data for easier retrieval.

- **Optimize Possibilities for Re-use:** By providing relevant keywords, the data becomes more discoverable and accessible to others who might want to reuse it for their own research or applications. This increases the value and impact of the data by broadening its potential audience and applications.

For keywords at this stage of the project the following list is proposed:

- Climate change adaptation
- Water harvesting
- Water scarcity management
- Water Framework Directive
- Decision framework, knowledge orchestrator, co-creation, adaptation pathways, alternative water resources,integrated water management

3.1.5 Approach for clear versioning

Zenodo repository and Argos (digital DMP) standardization will ensure that data is stored under specific structure to be easily identified on a historical basis.

3.1.6 Specify standards for metadata creation

Metadata standards will be required to:

(i) establish a common understanding of the meaning of the data, and

(ii) ensure correct and proper use and interpretation of the data by its owners and users.

To achieve this, a few characteristics, or attributes of the datasets will be defined and are described in the following section.

3.1.7 Type of metadata created and how

Along with the metadata described in Section 0 for the name of the datasets, metadata are also created for each dataset to describe the size of the file, its format, the data provider (owner), last update (date), time of update. In addition, a very short description of each dataset is also provided:

Table 9 Metadata Information

METADATA	BRIEF DESCRIPTION
Size (KB)	Corresponds to the size of the file.
Format	Indicates the format of the represented information. It could be JSON, CSV, etc.
Short Description	A brief explanation of the dataset.
Data Provider	Name of the company and/or person who provides the information.
Last Update (data)	Date of the last update performed over the dataset.
Time of Update	Timestamp (DD-MM-YYYY, HH:MM) in which update has been performed.

3.2 Making data openly accessible

RECREATE will generate public datasets available as Open Access-(OA). More specifically, RECREATE will release OA datasets in reference for the cost-efficiency, environmental and health impact of different technological and water management alternatives. All of these mentioned datasets will be uploaded into Zenodo, Argos and EOSC. Moreover, the output of the models applied to the case-studies will also published (OA) before the end of RECREATE. In this regard, permissions will be requested to the specific case-study partners and data providers.

This section will be updated on next iterations to provide detailed information on how data will be made accessible, assessable and intelligible. More specifically:

- Specifics on which data will be made openly available
- Which data is kept closed and provide the rationale?
- How the data will be made available
- What methods and software tools are used to access the data?
- Documentation of software needed to access the data included
- Inclusion of relevant software
- Data and associated metadata, documentation and code deposit
- Provision of access provided in case of restrictions

As detailed in Section 3, all data, information, and knowledge considered relevant for the scientific community will be made accessible under Open Access. When a dataset is set to be accessible publicly, this information will be fulfilled and the DMP updated accordingly.

3.2.1 Specifics on which data will be made openly available

In the following Table 10 we provide an indicative time schedule for all data publications will be provided in open access; however this table and its dates could change during the project:

Table 10 Tentative time plan for open access data publications

INDICATIVE TIMETABLE FOR DATA RELEASE IN OA	
Month 48	Scientific papers with projects outcome available in journals
Month 3-35	Testing and demonstration phase will be collect data from CSs, remote sensing, satellite information and users interactions and tool outcomes, in addition to weather data and sources for climate analysis

Month 1 - 48	Personal data collected in dissemination and communication phase
Month 30 - 42	information and main results of stakeholder engagement

As RECREATE is a multi-beneficiary project it is also possible for specific beneficiaries to keep their data closed according to relevant provisions in the consortium agreement.

3.2.2 Which data is kept closed and provide the rationale?

For the time being no data is considered as closed.

3.2.3 How the data will be made available?

Data will be made available through Zenodo Repository, which is compliant with Horizon Europe regulations. In general, for Public Availability of Data, data will be shared when the related deliverable, paper or data set has been made available at an Open Access (OA) repository from the responsible partner/owner of the data. It is expected that data related to a publication will be openly shared. However, to allow the exploitation of any opportunities arising from the raw data and tools, data sharing will proceed only if all co-authors of the related publication agree. The Lead author is responsible for getting approvals and then sharing the data and metadata on Zenodo. The Lead Author will also create an entry on Argos/OpenAIRE to link the publication to the data and the virtual version of the DMP. Argos/OpenAIRE is a service built to offer this functionality and may be used to reference both the publication and the data. A link to the Argos/OpenAIRE entry will then be submitted to the RECREATE Website Administrator by the Lead Author. As described above, pertinent approvals have been received from the data owners.

In view of the precautions for protection of personal data, it is explicitly confirmed that the data collected will be publicly available, after care is taken with regards to rules of confidentiality, anonymity, and protection. Anonymized final data sets will be open access and procedures are set as to how data will be preserved and archived in the repository. We are aware of post-publication risks to local researchers and end-users in our research sites and will mitigate all reasonable risk before publication according to the ethical and IPR requirements set.

However, “Opting Out” remains a choice for data owners, as it is possible that even though comprehensive measures are taken to ensure the safety of participants, researchers and their environment, it is only after a RECREATE report or peer reviewed article is published and generation of date sets is realized, that the question of open access arises. Open access does not entail an absolute obligation to publish all data, and it is up to researchers and associated organization to decide whether data is suitable and ethical to be published or not.

3.2.4 What methods and software needed to access the data included?

There are many technologies that can be exploited and adopted to perform this function.

The REST API service access is based upon the HTTP protocol with client and server requests and responses using XML or JSON. Client applications could use this interface for executing service repository queries and receiving service repository metadata results. The essential purpose of these tools is to enable the following general features:

- Locate, access and make use of resources in an open, distributed system by providing facilities for retrieving, storing and managing many kinds of resource descriptions.
- Model different datasets by considering different types of data.
- Access information from the different databases.

The REST API service managed by the Web of Things could store a multitude of resource descriptions as:

- Raw data from sensors
- GIS data
- Time-series
- End user and socio-economic information

Furthermore, to homogenize the interactions between the different information sources, a semantic layer and common data exchange model will be elaborated driven by an ontology to representing the information from multiple domains. The common data language will be encapsulated and integrated using JSON-LD representation.

Moreover, technology scouting about different database technologies will be performed to design and deploy a persistent layer for storing information about different data sources (entities) and their real-time information (time-series); potential candidates are:

- MongoDB and Fuseki for entities database
- InfluxDB for time-series database

3.2.5 Documentation of software needed to access the data included

The following tools and data repositories provide a sound documentation for accessing the data. (this part will be completed in next iterations)

3.2.6 Zenodo Repository

Zenodo is built and developed by researchers, in the context of The OpenAIRE project, that in the vanguard of the open access and open data movements in Europe, commissioned by the EC to support their nascent Open Data policy by providing a catchall repository for EC funded research. One of its mayor advantages is it works closely with GitHub, enabling users to make the work they share on GitHub citable by archiving one of your GitHub repositories and assigning a DOI with the data archiving tool Zenodo.

3.2.7 Argos Tool (Virtual Data Management Plan)

Argos tool is built and developed by the Open Aire community. Indeed, as explained in their website: "ARGOS is the joint effort of OpenAIRE and EUDAT to deliver an open platform for Data Management Planning that addresses FAIR and Open best practices and assumes no barriers for its use and adoption."[6]. Indeed, the tool permits to create a digital version of the data management plan to create actionable DMPs. Moreover, Argos services are directly connected to Open Aire Service Catalogue[7] and also EOSC catalogue[8]. Therefore, the uses or Argos will facilitate to publish open datasets and make it accessible to the research community openly across well-known communities and following FAIR principles.

3.2.8 Real-Time system

The real-time data is provided by an IoT sensor network (using standard communication protocols such as AMQP, CoAP, LoRaWAN) of the existing tools and technologies selected and improvides to provide data to RECREATE data lake. This data will be shared and accessible through the web by a REST API service based on open standards. Moreover, this data will be complemented with weather and climate data on different time horizons (short-term <= 15 days, long-term projection > 15 days) provided by cuttingedge data climate services (e.g. Copernicus Climate Change Service (C3S)) in order to build further knowledge (such as indicators/proxies) about climate impacts on water resources. The documentation of this IoT solutions will be provided under D1.3 (Harnessing digital technologies and tools for improved data collection and integration) where also will be described the use of digital twins and some AI models to improve the cost efficiency and the resilience of treatment systems.

[6] https://argos.openaire.eu/splash/about/how-it-works.html

[7] OpenAIRE Service Catalogue

[8] https://marketplace.eosc-portal.eu/service/openaire.argos

3.2.9 Inclusion of relevant software (e.g. in open source code)?

All existing software used (Argos, Zenodo, etc) is described in previous sections. However, RECREATE will require the development of some tools and data storage. The Table 11 represents the open-source frameworks and libraries that will be used in the development of the digital assets.

Table 11 Open Source libraries applied to RECREATE

OPEN SOURCE LIBRARIES	URL	RECREATE DIGITAL ASSET
Influx DB	https://www.influxdata.com/	A database to store time series efficiently.
Node.js	https://node.org/	Asynchronous event-driven JavaScript runtime environment, designed to build scalable network applications.
Angular Framework	https://angular.io/	Framework to develop different forms and dashboard to visualize information

3.2.10 Data and associated metadata, documentation, and code deposit

As stated in previous sections, the following tools are the ones associated with metadata, documentation, and code repository:

Table 12 RECREATE metadata, documentation and code repositories

REPOSITORY/TOOL	BRIEF DESCRIPTION
Zenodo	Zenodo metadata will serve to represent and link open research datasets.
Argos	A digital DMP and subsequent metadata will serve to link datasets and also, sustain open research in EU through Open Aire and EOSC.
Github	Code repository

(*) This table will be updated according to the project progress

3.2.11 Provision of access provided in case of restrictions

To ensure the FAIR management of the data, RECREATE_WT interoperable architecture and Zenodo repositories ensure an authorization scheme applied for accessing the data, depending on the scope of the usage. It is upon the partners to decide what is the most appropriate authorization scheme.

3.3 Making data interoperable

This section will be updated on next iterations of this Deliverable to provide detailed information on how data will be made interoperable to specific quality standards and more in detail:

- Assess the interoperability of project data
- Specifics on data/metadata vocabularies, standards, methodologies followed
- Use of standard vocabulary for all data types present to allow inter-disciplinary interoperability
- Provision of mapping to more commonly used ontologies

To ensure interoperability and homogenize the interactions between the different information sources, a semantic layer and common data exchange model will be elaborated driven by an ontology that allow representing information from multiple domains (developed as an extension of SAREF[9]). This ontology will be encapsulated an integrated inside RECREATE_WT REST API using JSON-LD representation. Complementing the REST-API, a real-time engine based on event-driven and reactive programming (using websockets) will be implemented to integrate and expose the information in real-time tool core-platform. Moreover, technology scouting about different database technologies will be performed to design and deploy a persistent layer for storing information about different data sources (entities) and their real-time information (time-series). Also the RECREATE_WT platform will be compatible with FIWARE using some of its modules in the design of the architecture.

3.3.1 Assess the interoperability of project data

All data collected and/or produced in the project will be interoperable, since they follow well documented international standards/formats, like XML, CSV, JSON, JSON-LD, HIS, INSPIRE[10] specification, OGC[11] specification, allowing data exchange and re-use between researchers, institutions, organizations, countries.

3.3.2 Specifics on data/metadata vocabularies, standards, methodologies followed

The initial description is provided in Section 4 of this Deliverable, and specifically in Table 9 and Table 12.

[9] SAREF

[10] https://inspire.ec.europa.eu/data-specifications/2892

[11] https://www.ogc.org/docs/is

3.3.3 Use of standard vocabulary for all data types present to allow inter-disciplinary interoperability

The initial description is provided in Section 3 of this Deliverable, and specifically in **Error! Reference s ource not found.**Table 9 and Table 12.

3.3.4 Provision of mapping to more commonly used ontologies

RECREATE knowledge repository is planned to provide their own ontology that will use terms from the following wide-known and public ontologies, such as, SAREF ontology. As a reference ontology for the representation of measures and properties (variables).

3.4 Increase data re-use (through clarifying licences)

This section will be updated on next iterations, to provide detailed information on how data will be made usable beyond the original purpose for which it was collected, and more in detail:

- Data licensing to permit the widest reuse possible
- Data availability for re-use
- Why and for what period a data embargo is induced
- Data useable by third parties after the end of the project
- Restriction of re-use of some data
- Data quality assurance processes
- Length of time for which the data will remain re-usable

As detailed in Section 3, all data, information, and knowledge considered relevant for the scientific community will be made accessible under Open Access. When a dataset is set to be accessible publicly, this information will be fulfilled and the DMP updated accordingly.

3.4.1 Data Licensing to permit the widest reuse possible

It is proposed to use one of this licenses as much as is possible and practicable, which allows sharing, remixing, transforming, and building upon the material for any purpose:

- Creative Commons Attribution International Public License (CC BY)
- Creative Commons Public Domain Dedication (CC 0)
- or a licence with equivalent rights

Products should be redistributed under the same license. It is not yet decided at this stage. It will be decided during development of RECREATE **Exploitation Plan** (Task 6.4, and 6.5).

3.4.2 Data availability for reuse

As envisioned in the Section 3.2.1, some data will be publicly available at M1-M48.

3.4.3 Why and for what period a data embargo is induced?

It is decided that all publications will be done until month 48 of the project and then data will be published through Green OA and Gold OA.

3.4.4 Data useable by third parties after the end of the project

Open Data used by third parties should be cited accordingly with the rules established by Zenodo/Argos in case of datasets and model outputs. These rules includes the name of the authors, the project, the DOI, the nature of the document and the year of publication:

AuthorSurname1, AuthorName1; AuthorSurname2, AuthorName2; …; AuthorSurnameN, AuthorNameN (year). Title_of_dataset [dataset]. Zenodo. Doi

3.4.5 Restriction of re-use of some data

There will be no restriction of use for the dataset published in open access. For the other data sources, if it's the case, it will be assessed with the consortium partners which rules to follow at M48.

3.4.6 Data quality assurance process

Data used for RECREATE will be derived from well-known complexity science models that will be constructed during the execution. These models to be elaborated will use information coming from available open datasets such as Copernicus, Meteocat, etc.

All these databases provide data with identified quality and provenance.

3.4.7 Length of time for which the data will remain re-usable

According to the objectives described in Section 2.1, EUT will host the tool in its own serverwill maintain it for 2 years after the project. Also, the maintenance of the sensors will be maintained for 2 years. Moreover, the records and support documentation will be maintained by the beneficiaries for a period of 5 years. After this period, the data will be maintained in Zenodo without any update.

4. Allocation of resources

This section is mainly devoted to the description of the associated costs of the FAIR data. Under this section, the following questions will be answered:

- What are the costs for making data FAIR in RECREATE?
- How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions).
- Who will be responsible for data management in RECREATE?
- Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

4.1 Costs for making data FAIR in RECREATE

The cost is estimated at 1-person month per case study to making data FAIR and maintainable. Thus, the total cost is estimated at 30.100€ including travel, other costs (7 person/months @4.300/p-m). This amount is already covered by the RECREATE project budget.

4.2 How will these be covered?

During the project life, costs are covered by the RECREATE budget. Following project closure, this cost will be covered by the pilots or new applications (impact analysis in WP6 of RECREATE GA) partners.

4.3 Who will be responsible for data management in your project?

EURECAT (EUT) will be responsible for the data management.

4.4 Resources for long term preservation

Long term preservation resources are:

- Costs: 30.100€ including travel, other costs (7 person/months @4.300/PM). This are the costs per year of maintaining datasets and the online digital assets.
- Potential value: updated for 2 years after the project's completion. After this timeframe, the value of the preserved database will be reevaluated. It depends on the Project's exploitation.
- Who decides and how: the RECREATE Consortium decides on the duration of the long-term preservation of the data.
- What data will be kept: All the data used for the RECREATE application in the case studies.
- For how long: The data will be preserved for 2 years after project completion. After this period, the data has no relevance, and unless the project is exploited with additional applications, the database will be outdated.

5. Data security

This section is devoted to the data security aspects to be implemented and considered within RECREATE. In this regard, the following questions will be answered:

- What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?
- Is the data safely stored in certified repositories for long term preservation and curation?

5.1 Provisions for data security (including data recovery as well as secure storage and transfer of sensitive data)?

All provisions for data security will be established by the following data repositories and digital assets:

- **Zenodo:** As wide used repository, have their own data securization, backup strategies and data accessibility and authorship.

5.2 Is the data safely stored in certified repositories for long term preservation and curation?

For the case of Zenodo and according to their policies on longevity, the following capacities will be adopted for data preservation and curation:

- **Versions:** Data files will be versioned. The uploaded data will be archived as a Submission Information Package. Derivatives of data files will be generated, but original content is never modified. Records can be retracted from public view; however, the data files and record are preserved.
- **Replicas:** All data files will be tentatively stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files will be kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- **Retention period:** Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.
- **Functional preservation:** Zenodo makes no promises of usability and understandability of deposited objects over time.
- **File preservation:** Data files and metadata are backed up nightly and replicated into multiple copies in the online system.

- **Fixity and authenticity:** All data files will be stored along with a MD5 checksum of the file content. Files are regularly checked against their checksums to assure that file content remains constant.
- **Succession plans:** In case of closure of the repository, best efforts will be made to integrate all content into suitable alternative institutional and/or subject based repositories.

6. Ethical aspects

This section is mainly devoted to the description of the Ethical Aspects adhered to the RECREATE project. Under this section two main aspects will be considered, the Intellectual Property Rights and the assessment and protection of personal data and research activities information under the framework of the GDPR and compatible regulations.

6.1 General

Within RECREATE, only general ethical issues are concerned such as informed consent, anonymity and confidentiality associated with the voluntary involvement of human participants in the European Union. Types of such data collected in RECREATE are user interviews, opinions and reviews associated with project's components. Non-exhaustive list is as follows:

- Stored involvement of RECREATE self-learning assessment engine users to gain insight into the decisions and behaviours of the stakeholders and to allow further analysis for improve decision-making.
- The graphical use interface - to collect information from users so that the self-learning assessment engine can learn from user decisions.
- A series of interviews with stakeholders and decision makers.
- Planned contacts with representative persons of targeted users. Interviews should be carried out by phone/online or face-to-face when convenient. Interviews should help define the expected functionalities/ services to be offered, test the price that could be acceptable and identify distribution channels to access these clients
- The end-users, potential developers, and partners, etc. will be provided the opportunity to test and review the latest products and services.
- Methodology and procedures for sensitive data processing and storing will be specified as a part of the ethics. It is important to emphasize that special efforts will be devoted to anonymizing information and securing accessibility. Mechanisms to delete personal data will be provided in an easy and usable manner.

To strengthen further commitment of RECREATE partnership research, this approach follows good ethical practices will ensure fair and equal power relationships between researchers and participants, the consortium agrees to comply with the principles laid down in the European Code of Conduct for Research Integrity, published by the European Science Foundation [1]. These principles include:

- honesty in communication of the research's goals and intentions, in reporting methods and procedures and in conveying interpretations.
- reliability in performing research.
- objectivity, which requires facts capable of proof, and transparency in the handling of information.
- impartiality and independence.

- openness and accessibility.
- duty of care - all researchers have a duty of care for the humans, animals, the environment, or the objects that they study.
- fairness in providing references and giving credit for the work of others.
- responsibility for the scientists and researchers of the future.
- care will be taken to minimize the potential collection of personal data, i.e. while taking photos and/or videos during events.

In this regard, RECREATE will not involve any potentially vulnerable groups or people unable to consent (children, those with a learning disability or cognitive impairment, or individuals in a dependent or unequal relationship), and it will not involve sensitive topics which might induce psychological stress, anxiety or humiliation, deception, or any potential increased danger to participants, or the collection of personal data from participants.

Further, it will not involve the collection or processing of the following types of data:

- Research involving sensitive topics - for example participants' sexual behaviour, their illegal or political behaviour, their experience of violence, their abuse or exploitation, their mental health, or their gender or ethnic status;
- Research involving groups where permission of a gatekeeper is normally required for initial access to members - for example, ethnic or cultural groups, native peoples or indigenous communities;
- Research involving deception, or which is conducted without participants' full and informed consent at the time the study is carried out;
- Research involving access to records of personal or confidential information, including genetic or other biological information, concerning identifiable individuals;
- Research which would induce psychological stress, anxiety or humiliation or cause more than minimal pain;
- Research involving intrusive interventions - for example, the administration of drugs or other substances, vigorous physical exercise, or techniques such as hypnotherapy. Participants would not encounter such interventions, which may cause them to reveal information, which causes concern, in the course of their everyday life;
- Research involving the tracking or observation of participants (e.g. surveillance or localization data, and Wide Areas Network -WAN- data, such as IP address, MACs, etc.). However, ‘cookies’ will used in the website and the graphic user interface to help analyse how users behave while interact with the RECREATE platform.
- A privacy statement will be put on the website regarding the use of external services like Google Analytical (or similar) to track and get statistics from users in the use and interaction with the website. A similar privacy statement will be put on the graphical user interface with

similar purposes. Moreover, it is important to notice that none of the data collected by the RECREATE project requires a notification or authorisation for the collection and/or processing of the personal data to authorities or other responsible entities.

In order to ensure that the RECREATE partnership's participatory research approach follows good ethical practice and ensures fair and equal power relationships between researchers and participants, the consortium will all agreed that they will sign, make public and implement an ethics agreement, based on the European Code of Conduct for Research Integrity, published by the European Science Foundation [2].

Moreover, RECREATE consortium also agrees to follow up the rules and guidelines of the GDPR EU regulation adhered to the data privacy and security of personal information across their digital and non-digital developments[12]. In this regards, the following conditions enables to elaborate the Data Impact Assessment[13] to ensure correct protection of the users information:

- Using and elaboration of newer technology
- Track of people's location and behaviour.
- Systematically monitoring a public accessible place on a large scale.
- Processing personal data related to "racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation".
- Data processing is used to make automated decisions about people that could have legal (or similarly significant) effects.
- Processing children's data.
- Processing could result in physical harm to the data subjects if it is leaked.

6.2 Intellectual Property Rights (IPR)

In RECREATE, the IPR management strategy will be developed in the Task 7.3 and the detailed results will be presented in D7.3, to create a favourable environment for further exploitation of the generated results and knowledge (refer 0). All information regarding the IPR management will be clearly defined in the 'Consortium Agreement' (CA). Throughout, our strategy for the management of IPR will follow accepted guidelines on authorship (using the guidance provided by the IPR Helpdesk) and standard forms of copyright for academic publications, including green and gold routes for open access, while making other project outputs open access using Creative Commons licensing.

Considering the nature of the project and expected high TRLs at its completion, the future commercial exploitation opportunities of the research and innovation generated in RECREATE could

12 https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN
13 https://gdpr.eu/data-protection-impact-assessment-template/

be significant. A mix of public and private contributions will generate a balanced, inclusive, and resilient development, while ensuring that products and services are affordable and accessible to all. We will maximise opportunities to develop both for profit and not-forprofit exploitation of our research and outputs. We will follow standard industry approaches that recognise IPR protection measures (e.g., patents, design rights, copyright, trade secrets, etc.). The mechanisms to handle IPR management will be based on a three-level approach:

(i) **Proposal phase IPR:** All partners' background assets (i.e., scientific studies, methods, tools) and potential intellectual property rights (IPR) attached to them (i.e., patent, copyright) likely needed for the implementation of the project and/or for the use of the expected results that can be subject to IPR, have been considered during the proposal preparation. In addition, the consortium has carried out a prior SOTA analysis (refer 1.1.3 of GA) to avoid any future IPR conflicts. Specifically, preliminary searches were performed, using open resources (Google Patents, Patent Lens) and tools provided by EUIPO (Espacenet, eSearch plus), EUIPN (TMview) and WIPO (Global Brand Database). The analysis revealed that there are no trademarks on RECREATE's acronym, logo, and its full title.

(ii) **IPR during the project:** The Horizon IP Scan service will be utilised at the project's Grant preparation phase to carry out a preliminary assessment of intangible assets with the aim to identify any IPR issues that the consortium will potentially encounter. During the project, the newly generated knowledge and IPR will be recorded, recognised, captured, and assessed. Protective and supportive measures, such as the obligation to sign Non-Disclosure Agreements with external stakeholders / contacts, will be undertaken by the consortium to create confidence to the involved participants, ensuring a transparent allocation of IPR. In addition, the consortium plans to openly publish the overall project results and all publications will be stored in trusted repositories (i.e., Zenodo). A Result Ownership List (ROL) will be defined among partners and provided with the final periodic report.

(iii) **Post-project IPR:** Confidentiality obligations, transfer of results, obligations to protect results capable of commercial exploitation will also be considered after the end of the project. A Freedom-to-Operate (FTO) analysis has been carried out which resulted in no barriers, and therefore RECREATE pre-identified and newly generated IPRs can and will be protected. Results non-exploited by owners within one year after the end of the project will be published on Horizon Results Platform for potential exploitation by another legal entity (e.g., through transfer or licensing).

7. Other issues

At the moment, no other procedures for data management are envisioned. If there are some changes on this, we will update the DMP accordingly in the D7.3 (due in M24) and in D7.5 (due in M48).

8. Conclusions & Future Work

This section is mainly devoted to the description of the main conclusions of the elaboration of this initial version of the DMP. Furthermore, this section describes future work in relation to the DMP their implicit datasets, ethics and data security.

8.1 Conclusions

The present deliverable has been mainly focused on the elaboration of the initial version of the data management plan of RECREATE. In this regard, this document have described main strategies and methodologies for the publication open research data. Moreover, the present document has tackled the possibility of having close proprietary data and information that permits to evolve RECREATE digital tools.

Complementary to the identification of open data information, the document has been focused on the management of the information according to make it compatible and ensure the FAIR principles. Moreover, the present document also tacked the ethics and privacy aspects of the project.

As a main conclusion, the present document establishes the basis for a correct management of data across the entire project. A part of this, the project has established a methodology and guidelines to ensure data privacy an ethics. Another important aspect is the elaboration of a digital version of the DMP (in parallel to this document) just to make it compatible with the FAIR principles and share datasets across main open repositories as Zenodo, Open Aire and EOSC.

8.2 Future Work

As remarked within the document, the DMP is a live document that will be continuously updated within the project. In this regard, the future envisioned actions for the next version (M42) are:

Table 13 Future Work Actions in relation to the DMP

FUTURE WORK ACTIONS	DESCRIPTION
Update of the datasets list	Update of the datasets as the evolvement of the demo-cases interaction and also, the development of the RECREATE digital platform and physical assets.

Maintenance of the virtual version of the DMP	Publication of open RECREATE datasets in the Argos platform in order to link it with Zenodo and make it available for further research.
Data privacy and Ethics	Ensure the user accessible digital assets (websites, mobile APPs, exploration tools, etc) accomplish the EU regulations in data privacy and ethics.

9. References

[1] European Science Foundation, European Code of Conduct for Research Integrity, 2017.
[2] European Science Foundation, RepoCode Of Conduct for Research Activities, 2011.

In case of any questions, please contact:

Eloy Hernandez
Senior Project Manager & WP3 lead

eloy.hernandez@eurecat.org

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101136598. This document reflects only the views of RECREATE consortium, neither the European Commission nor any associated parties are responsible for any use that may be made of the information it contains.