D11.4 – Open Research Data Pilot and Data management Plan

Project Information

Grant Agreement Number	958454
Project Full Title	Intelligent Water Treatment for water preservation combined with simultaneous energy production and material recovery in energy intensive industries
Project Acronym	intelWATT
Funding scheme	IA
Start date of the project	1st October 2020
Duration	42 months
Project Coordinator	Andreas Sapalidis (NCSR)
Project Website	https://www.intelwatt.eu

Deliverable Information

Deliverable n°	D11.4
Deliverable title	Open Research Data Pilot and Data management Plan
WP no.	11
WP Leader	NCSR
Contributing Partners	ALL
Nature	ORDP: Open Research Data Pilot
Authors	WG
Contributors	ALL
Reviewers	
Contractual Deadline	31/03/2021
Delivery date to EC	31/03/2021

Dissemination Level

PU	Public	✓
PP	Restricted to other programme participants (incl. Commission Services)	
RE	Restricted to a group specified by the consortium (incl. Commission Services)	
CO	Confidential, only for the members of the consortium (incl. Commission Services)	

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

Document Log

Version	Date	Author	Description of Change
V1.0	29/3/2021	M. Rinaldi	First release
V2.0	31/3/2021	A. Sapalidis	Final document

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

Table of Contents

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

1 Executive Summary

The purpose of the **Open Research Data Pilot and Data Management Plan** (DMP) is to illustrate data management policy that is and will be used by the beneficiaries with regard to all the data sets that are collected, processed and/or generated by the project.

The final objective of data management in the context of a Horizon 2020 project is to make research data **findable, accessible, interoperable and reusable (FAIR)**, with the aim to improve access to scientific information and boost the benefit of the public investments in research.

Therefore, in this document the project approach to open access and management of research data is addressed considering:

- the handling of research data during and after the end of the project
- what data will be collected, processed and/or generated
- which methodology and standards will be applied
- whether data will be shared/made open access
- how data will be curated and preserved (including after the end of the project where relevant).

This document is the first version of the DMP. The DMP will be updated over the course of the project whenever significant changes arise, such as (but not limited to):

- new sets of data,
- changes in consortium policies (e.g. new innovation or exploitation potential),
- changes in consortium composition and external factors (e.g. new consortium members joining or old members leaving).

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

2 Introduction

The different research datasets collected, generated and analysed within the project activities have been profiled considering available data in this phase, in order to understand the possibility of making them FAIR (findable, accessible, interoperable and reusable). Data sets thus identified are illustrated in the following section.

As the project receives funding from the European Union, the consortium in any case commits to respect both the Grant Agreement requirements and the overall EU Policies with respect to giving broader access to scientific publications and data to allow the capitalisation of previous research results, with the ultimate aim and consequence of accelerating innovation.

Therefore, non-sensitive publishable results will be possibly shared and made accessible to the research community for verification, re-use and cross-validation scientific purposes, preferably addressing journal articles as the dominant type of peer-reviewed scientific publication, focusing on high impact factor journals, and then applying to Gold or Green Open Access.

3 Consortium Agreement Provisions

Provisions of the Consortium Agreement, in line with the Grant Agreement, have to be considered in the Management of Research Data generated within the project.

The Ownership of Results is ruled by the provisions of the Grant Agreement, Article II.26. - Article II.31. Moreover, the Consortium Agreement provides specific indications about Joint Ownership, as follows:

> ***8.1. Ownership of Results***
>
> *Results are owned by the Party that generates them.*
>
> ***8.2. Joint ownership***
>
> *Joint ownership is governed by Grant Agreement Article 26.2 with the following additions: Unless otherwise agreed:*
>
> *- each of the joint owners shall be entitled to use their jointly owned Results for non-commercial research activities on a royalty-free basis, and without requiring the prior consent of the other joint owner(s), and*
>
> *- each of the joint owners shall be entitled to otherwise Exploit the jointly owned Results and to grant non-exclusive licenses to third parties (without any right to sub-license), if the other joint owners are given:*
>
> *(a) at least 45 calendar days advance notice; and*
>
> *(b) Fair and Reasonable compensation.*

Considering publication as the first step of making data open, the project Consortium Agreement - in line with Article 29.1 of the Grant Agreement - provides a standard procedure to ensure that publications do not infringe exploitation potential and commercial interests of any results owner, specifically:

> ***8.4.2. Dissemination of own Results***
>
> ***8.4.2.1.***
>
> *During the Project and for a period of 1 year after the end of the Project, the dissemination of own Results by one or several Parties including but not restricted to publications and presentations, shall*

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

be governed by the procedure of Article 29.1 of the Grant Agreement subject to the following provisions. Prior notice of any planned publication shall be given to the other Parties at least 45 calendar days before the publication. Any objection to the planned publication shall be made in accordance with the Grant Agreement in writing to the Coordinator and to the Party or Parties proposing the dissemination within 30 calendar days after receipt of the notice. If no objection is made within the time limit stated above, the publication is permitted.

8.4.2.2.

An objection is justified if

(a) the protection of the objecting Party's Results or Background would be adversely affected (b) the objecting Party's legitimate interests in relation to the Results or Background would be significantly harmed.

(c) the proposed publication includes the Confidential Information of the objecting Party The objection has to include a precise request for necessary modifications.

8.4.2.3.

If an objection has been raised the involved Parties shall discuss how to overcome the justified grounds for the objection on a timely basis (for example by amendment to the planned publication and/or by protecting information before publication) and the objecting Party shall not unreasonably continue the opposition if appropriate measures are taken following the discussion.

8.5.

The objecting Party can request a publication delay of not more than 45 calendar days from the time it raises such an objection. After 45 calendar days the publication is permitted.

8.5.1. Dissemination of another Party's unpublished Results or Background

A Party shall not include in any dissemination activity another Party's Results or Background without obtaining the owning Party's prior written approval, unless they are already published.

8.5.2. Cooperation obligations

The Parties undertake to cooperate to allow the timely submission, examination, publication and defence of any dissertation or thesis for a degree that includes their Results or Background

Also, Access Rights to Background and Results - and therefore to related research data - are ruled:

9.2. General Principles

9.2.1.

Each Party shall implement its tasks in accordance with the Consortium Plan and shall bear sole responsibility for ensuring that its acts within the Project do not knowingly infringe third party property rights.

9.2.2.

Any Access Rights granted expressly exclude any rights to sublicense unless expressly stated otherwise.

9.2.3.

Access Rights shall be free of any administrative transfer costs.

9.2.4.

Access Rights are granted on a non-exclusive basis.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

9.2.5.

Results and Background shall be used only for the purposes for which Access Rights to it have been granted.

9.2.6.

All requests for Access Rights shall be made in writing. The granting of Access Rights may be made conditional on the acceptance of specific conditions aimed at ensuring that these rights will be used only for the intended purpose and that appropriate confidentiality obligations are in place.

9.2.7.

The requesting Party must show that the Access Rights are Needed.

[…]

9.4.

Access Rights for Exploitation

9.4.1.

Access Rights to Results if Needed for Exploitation of a Party's own Results shall be granted on Fair and Reasonable conditions. Access rights to Results for internal research activities shall be granted on a royalty-free basis.

9.4.2.

Access Rights to Background if Needed for Exploitation of a Party's own Results, including for research on behalf of a third party, shall be granted on Fair and Responsible conditions.

9.4.3.

A request for Access Rights may be made up to twelve months after the end of the Project

Additional Access rights - to external parties besides the ones for the partners part of the Consortium and the affiliated entities indicated in annex to the Consortium Agreement - can also be granted, upon specific agreements:

9.6. Additional Access Rights

For the avoidance of doubt any grant of Access Rights not covered by the Grant Agreement or this Consortium Agreement shall be at the absolute discretion of the owning Party and subject to such terms and conditions as may be agreed between the owning and receiving Parties.

4 The DATA Approach

The project generates a huge amount of data mainly coming from the experimental validation activities but also dealing with theoretical and LCA activities in this context the following main data sets groups can be identified:

DATA SET n. 1 – Experimental– WP2,3,4,5,7 – Contributing PARTNER: NCSR

DATA SET n. 2 – Experimental – WP2,3,4,9,10,11 – Contributing PARTNER: CNRS

DATA SET n. 3 – Experimental– WP 2,3,4,6 – Contributing PARTNER: CNR

DATA SET n. 4 – Experimental – WP2,5 – Contributing PARTNER: PPC

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 5 – Experimental – WP2,3,4,7 – Contributing PARTNER: THK

DATA SET n. 6 – Experimental – WP3 – Contributing PARTNER: UOB

DATA SET n. 7 - Experimental – WP2,3,4,8 – Contributing PARTNER: POLITO

DATA SET n. 8 – Experimental – WP3,4,5,6,7 – Contributing PARTNER: CUT

DATA SET n. 9 – Experimental – WP2,3,4,7 – Contributing PARTNER: BIA

DATA SET n. 10 – Experimental – WP2,5 – Contributing PARTNER: IHE DELFT

DATA SET n. 11 – LCSA Data – WP9 – Contributing PARTNER: STUDIO FIESCHI

DATA SET n. 12 – Theoretical – WP2,3,4,5,6,8 – Contributing PARTNER: TECHEDGE

DATA SET n. 13 – Experimental – WP2,3,6 – Contributing PARTNER: ACSA

DATA SET n. 14 – Replication study – WP9 – Contributing PARTNER: UJ

DATA SET n. 15 – Experimental – WP2,3,4,6,8 – Contributing PARTNER: REDSTACK

DATA SET n. 16 – Experimental – WP3,6 – Contributing PARTNER: CIEMAT

The different research datasets are and will be handled in a specific way, considering the intrinsic characteristics and level of confidentiality due to the relation to potentially exploitable results, as well as the entities data and information management policies.

Indeed, the expected exploitable results of IntelWATT include processes, smart sensing & monitoring and derived products in membranes and resins for water treatment,

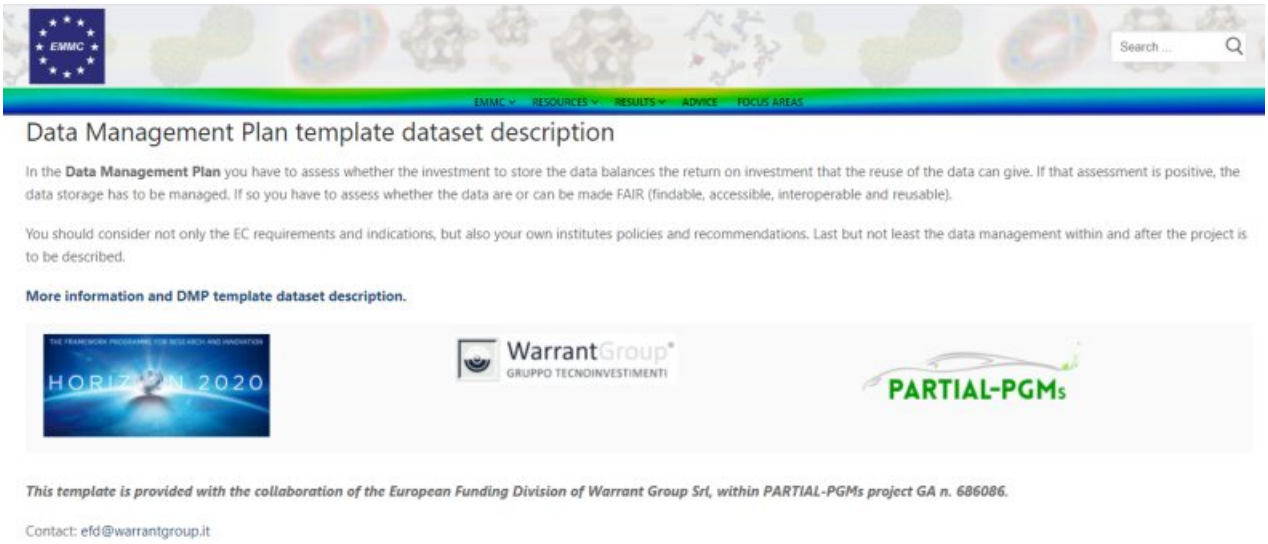
In this framework, the only data that are considered partially disclosable and therefore re-usable by Third Parties external to the Consortium on the long-term, both for the higher level of standardization and of interoperability, as well as for the lower level of confidentiality, are the one related to multiscale modelling and to protocols.

4.1 Data management plan (DMP) by specific datasets

A dataset template has been defined by Warrant Hub and approved by the European Materials Modelling Council (EMCC) that now suggest this approach to be used in Horizon 2020 Projects, the reference template can be downloaded at the following link:

https://emmc.info/emmc-info-data-management-plan-template-dataset-description/

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".



The definition of Research Data Sets is the basis for the definition of the Data Management Plan.

The idea is to take a “picture” of the research data generation and management within the project, and to assess if these data are or can be made FAIR (findable, accessible, interoperable and reusable), considering not only the EC requirements and indications, but also your own institutes policies and recommendations.

Tables were filled for each type of dataset (considering different types of data, with different formats, purposes, etc.).

“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 958454”.

4.1.1 DATA SET n. 1 – Experimental – WP2,3,4,5,7 – PARTNER: NCSR

DATA SET n. 1 – Experimental– WP2,3,4,5,7 – PARTNER: NCSR		
1 DATA SUMMARY	**Purpose of the Data**	Data concerns physicochemical analyses of relevant water streams for the case study No1 (as indicated and supplied by Partner No4, Public Power Company, Greece), optimization of hydrophobic membrane preparation parameters for membrane distillation, material characterization (such as ion exchange resins or activated carbon), membrane module design development and performance evaluation studies (relative to process conditions, technologies combinations, modules configurations and setups, membrane shape and dimension etc.) at the lab and pilot scale, process design drawing(s). All the generated data are in close relation with the implementation of the project's deliverables and objectives.
	Type and Format of data	▪ Text: field or laboratory notes, survey responses – in plain text, (txt), HTML, XLM, PDF/A {performance evaluation tests and material preparation} ▪ Numeric: tables, counts, measurements – in .XLSX, .CSV (performance evaluation tests) ▪ Audiovisual: images, video – in .JPEG, .JPG, .PNG, .TIFF, AIFF, WAVE, .MP3, .MP4 (microscopic techniques SEM, AFM) ▪ Instrument specific: equipment outputs (specify equipment and format) {physicochemical characterization and analysis - can be converted in XLS, ASCII and CSV format}
	Reused-Data	The majority of the data will be new, generated during the implementation of the project.
	Data origin	▪ Observational – data captured in real time (performance evaluation tests, material preparation), usually irreplaceable - i.e. sensor readings, images, telemetries, sample data. The data are reproducible if the same conditions and equipment are used. ▪ Experimental – data from lab equipment (physicochemical analysis, materials characterization), often reproducible - i.e. chromatograms, absorbance spectra, x-ray diffraction data, porometer, porosimeter. The observational and the experimental data are subject to analysis. In some case the analysis is very time consuming which increases the cost.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 1 – Experimental– WP2,3,4,5,7 – PARTNER: NCSR		
	Data size	Physicochemical analyses: - Revisable - Data file size depending on the analytical technique and instrument. The Datasets vary from some KBytes (AAS) to few MBytes (Ion chromatography) per experiment. Material characterization: - Revisable - Collection of data files from various laboratory instruments like SEM, AFM, contact angle, surface tension etc. Data files size depends on the instrument and (regarding microscopies) from the resolution of the images. Estimated file sizes from some KB to few MB (images) per measurement. Performance evaluation tests: - Revisable - Files of raw data acquired by the SCADA systems of experimental devises, controllers, online sensors and analyzers etc. ASCII files (in csv format) of approximately 10 MB per study depending on testing duration and sampling rate with an overall estimated size 50 GB . - The performance evaluation provides data that can be amended. In this case the test should be repeated in order to obtain new data.
	Data Security and Storage	Supplementary and raw data initially handwritten as text in lab books are not digitized but the information is transferred to digital reports stored in office computers. Experimental data concerning physicochemical analyses of water samples or characterization of materials (e.g. membranes) are stored in the hard drives of the corresponding instrument(s). Raw data collected during performance evaluation tests by SCADA systems of experimental devices are stored as .csv or spreadsheet (excel) files in lab computers. Reports, documents, photos, designs and other digitized material generated during – and relative to - intelWATT project are stored in office computers as well as in EmDESK platform. According to the lab policy all data are backed up regularly and the backups are stored in an internal network drive as well as removable hard drives. All the computers are protected by hard passwords, firewalls and antivirus / antimalware software. The current protocols prevent access to any external user.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 1 – Experimental– WP2,3,4,5,7 – PARTNER: NCSR		
	Data utility	The data are useful for the whole team of NCSR Demokritos, to all members of the intelWATT project as well as to other research groups focusing in membrane science/technology and relevant applications. End users in the electric power generation and renewable energy fields, water and wastewater treatment professionals, process engineers and post graduate students focusing in the activities of intelWATT are indicative additional groups which benefit from the project's results. The data can be further exploited as source for new funding or new internal scientific projects as well as for commercialization purposes.
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Discoverability of the data is achieved by making the analyzed data available (provided that rules and provisions of signed GA, CA and IPR interests are not violated) for publication in scientific journals. Discoverability is further maximized linking the article with the relevant dataset (including related metadata) which is deposited in a scientific data repository. These data can be further exploited as source for new funding or new internal scientific projects. The relative metadata are stored mainly in .txt, and .pdf format. Experimental keywords will be used to create the corresponding metadata. These will include some common fields like the name/code of the material (membrane or IX resin), partner's acronym, as well as some specific info depending on the technique. For example, for membrane evaluation the following fields will be included: Type of experiment, Feed composition, Temperature, Pressure, Flow rates, Sample info, etc. Even though the metadata file is not created by a standard schema and is in the form of a simple .txt file, it should provide the following information: Title - Name of the dataset or research project that produced it Creator - Names and addresses of the organization or people who created the data Identifier - Number used to identify the data Subject - Keywords or phrases describing the subject or content of the data Access information - Where and how your data can be accessed by other researchers Date - Key dates associated with the data, including time period covered by the data Methodology - How the data was generated, including equipment or software used, experimental protocol, other things one might include in a lab notebook Data processing - Information on how the data has been altered or processed

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 1 – Experimental– WP2,3,4,5,7 – PARTNER: NCSR		
		Sources - Citations to material for data derived from other sources, including details of where the source data is held and how it was accessed List of file names - List of all data files associated with the dataset, with their names and file extensions File Formats - Format(s) of the data, e.g. FITS, SPSS, HTML, JPEG, and any software required to read the data File structure - Organization of the data file(s) and the layout of the variables, when applicable Variable list - List and description of variables in the data files, when applicable Code lists - Explanation of codes or abbreviations used in either the file names or the variables in the data files (e.g. '999 indicates a missing value in the data')
	Identifiability of data (refer to standard id mechanisms)	Data is made identifiable by manually naming data files according to the rules of the lab and the file naming conventions referred below. No standard ID mechanism is used. Data (in the form of document, spreadsheets, images, presentation etc.) uploaded to an online data repository is identified by a Digital Object Identification (DOI) which assigned automatically by the repository. The DOI also contains metadata that provides the relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 1 – Experimental– WP2,3,4,5,7 – PARTNER: NCSR		
	Naming conventions used	All the results and data generated during the implementation period of the project will be stored in dedicated folders. The folders' structure is tree like. The root folder (named with the project acronym "intelWATT") contains the WP folders. Further structuring inside the WP folders will be based on Task or Case study numbering. For example folder ..\\intelWATT\WP_2\TASK_3 contains files and data relative to the task 3 of WP 2. Data files are manually named in a way that their source can be uniquely distinguished. For example, a name consisting of a prefix (e.g. characteristic of the experimental technique or material) followed by a date and/or time stamp could be an appropriate convention for a file containing raw data from an experimental device's SCADA systems. In addition, the raw data files should be linked with relevant records (in physical form in a lab book or in a text file) containing all the necessary details for the reusability / interoperability of the data (e.g. the experimental set up and conditions, the preparation method of the materials, pre and / or post treatments of the samples etc.). <u>*Second option*</u> *Data are manually named in a way that one can distinguish their source. Also the name of the interpreter/analyser is added.* *These data are used for the preparation of internal reports. The reports are named by the project and the period, in which the data belongs.*
	Search keywords approach	In some cases some keywords are used including the name of the sample and the corresponding technique used. <u>*Second option*</u> *No specific approach is provided. The pattern for keywords generation is mainly based on the kind of experiments performed, the experimental procedures and conditions, characteristics and properties of the under testing materials, analytical methods etc. As an example, for the performance evaluation of a PVDF membrane the recommended keywords are: PVDF membranes, flat sheet modules, membrane distillation, hydrophobicity, heat recovery etc.*
	Clear versioning approach	Revised datasets will take the extension "rev" followed by the number of the revision made. <u>*Second option*</u> *The versioning of data files is manual indicating the date of its creation and the writer.*

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 1 – Experimental– WP2,3,4,5,7 – PARTNER: NCSR		
	Standards or procedures for metadata creation applied	The option is between N/A and the text below Specific templates will be prepared for the collection of all the necessary information related to a particular dataset. A supplementary .pdf or .txt file will be created using the above information and a standardized layout / format. This file will be uploaded in a data repository where a DOI will be automatically assigned to it. The DOI contains metadata that provides users with relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored. Even though the project discipline is not related directly to any metadata standard or schema, however there are some general purpose standards like the Dublin Core Metadata Initiative that can be applied. Metadata Editors (preferably open access) will be used to generate partly or complete the relevant metadata files
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	N/A
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	The raw data obtained by the NCSR team belongs to the NCSR. The research team will define the way and the time that these data can be presented. The analyzed data that are presented to the consortium are available for the whole consortium for the full duration of the project and according to the rules and provisions of the signed G.A., C.A. and intellectual property obligations. Accessibility of data out of the scope of intelWATT project requires the permission of NCSR. The same occurs also for the raw data. In such a case the raw data can be made available only upon request and under separate agreement between the partners and as long this request does not interfere with future publication and intellectual property interests or precede data analysis.
	How data will be made available	Data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Disclosable data will be deposited in an open data repository, such as Zenodo or other recommended repositories (see below) provided that rules and provisions in signed GA and CA, IPR obligations or/and confidentiality restrictions are not violated. An option being considered is to link the data and metadata to forthcoming publications. Specific data can be made available upon requests (see 2.4)
	Methods or SW tools for data access	Tools for accessing the data are any text editor and spreadsheet packages for the output files. Optionally photo/image viewers / editors or CAD software may be needed for the visualization of some engineering designs.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 1 – Experimental– WP2,3,4,5,7 – PARTNER: NCSR		
	SW documentation and other information needed	In some cases, the supplementary notes may be needed for better understanding of the data.
	Repository for deposit of data, metadata, documentation and code	Specific repositories that will be considered include: FigShare, Mendeley Data, Open Science Framework, Harvard Dataverse, Dryad Digital Repository, Science Data Bank, Zenodo, which can be used for storage of data and metadata. These generalist repositories may also be appropriate for archiving associated analyses, or experimental-control data. *Second Option* *N/A*
	Access restrictions	As a rule, the availability of the data is according to the rule and the provisions of the signed G.A. and C.A., especially to the IPR clauses indicated there. In some circumstances, NCSR will not be able to disclose some of this data until a certain time within the project timeframe due to possible commercial confidentiality and intellectual property reasons.
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A
	Standard vocabulary or mapping to commonly used ontologies	N/A
	Information Portability and exporting	N/A
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)	**Data licensing for wide reuse**	N/A
	Timing of data availability for re-use (incl. indications on embargo)	Data can be made available after publication in scientific papers. The raw data will be not available to other partners, but can be shared only after permission of the NCSR Demokritos team.
	Data usability by Third Parties (after the end of the project)	Except for the case where results have been published, data could be reused only with the permission of NCSR and according to the rules established in the G.A. and C.A.
	Restrictions to data re-use	Restrictions to the re-use of data from the Consortium Partners as imposed by the signed G.A. and C.A. rules, specific IPR established agreements and/or possible commercial confidentialities

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 1 – Experimental– WP2,3,4,5,7 – PARTNER: NCSR		
	Quality assurance process	Data quality is assured through the : - Regular checking and maintenance of the experimental and mechanical equipment including electronic sensors and control devices. - Calibration of sensors (temperature, pressure, flow, conductivity etc.) - Calibration and performance verification of analytical instruments - Use of standards for certain measurements - Blank experiments and repetition of experiments. - Data entry validation and peer review of data.
	Length of time of data re-usability	N/A
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	N/A
	Data Management Responsibilities	The NCSR team is responsible for the management of the data obtained within the frame of intelWATT project. Researchers of NCSR are high educated and qualified to ensure that dataset is accurate and reliable.
	Long Term Preservation assessment	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

4.1.2 DATA SET n. 2 – Experimental – WP2,3,4,9,10,11 – PARTNER: CNRS

DATA SET n. 2– Experimental – WP2,3,4,9,10,11 – PARTNER: CNRS		
1 DATA SUMMARY	**Purpose of the Data**	Data concerns optimization of membrane preparation parameters for RO distillation membranes, material characterization (SEM, AFM, contact angle, FTIR), and performance evaluation studies (relative to process conditions, technologies combinations, modules configurations and setups, membrane shape and dimension etc.) at the lab and pilot scale, process design drawing(s). All the generated data are in close relation with the implementation of the project's deliverables and objectives.
	Type and Format of data	▪ Text: field or laboratory notes, survey responses – in plain text, (txt), HTML, XLM, PDF/A {performance evaluation tests and material preparation} ▪ Numeric: tables, counts, measurements – in .XLSX, .CSV (performance evaluation tests) ▪ Audiovisual: images, video – in .JPEG, .JPG, .PNG, .TIFF, AIFF, WAVE, .MP3, .MP4 (microscopic techniques SEM, AFM) ▪ Instrument specific: equipment outputs (specify equipment and format) {physicochemical characterization and analysis - can be converted in XLS, ASCII and CSV format}
	Reused-Data	The majority of the data will be new, generated during the implementation of the project.
	Data origin	▪ Observational – data captured in real time (performance evaluation tests, material preparation), usually irreplaceable - i.e. sensor readings, images, telemetries, sample data. The data are reproducible if the same conditions and equipment are used. ▪ Experimental – data from lab equipment (physicochemical analysis, materials characterization), often reproducible - i.e. chromatograms, absorbance spectra, x-ray diffraction data, porometer, porosimeter. The observational and the experimental data are subject to analysis. In some case the analysis is very time consuming which increases the cost.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 2– Experimental – WP2,3,4,9,10,11 – PARTNER: CNRS		
	Data size	Physicochemical analyses: - Revisable - Data file size depending on the analytical technique and instrument. The Datasets vary from some KBytes (DLS) to few MBytes (microscopy) per experiment. Material characterization: - Revisable - Collection of data files from various laboratory instruments like SEM, AFM, contact angle, surface tension etc. Data files size depends on the instrument and (regarding microscopies) from the resolution of the images. Estimated file sizes from some KB to few MB (images) per measurement. Performance evaluation tests: - Revisable - Files of raw data acquired by the SCADA systems of experimental devises, controllers, online sensors and analyzers etc. ASCII files (in csv format) of approximately 10 MB per study depending on testing duration and sampling rate with an overall estimated size 50 Gb. - The performance evaluation provides data that can be amended. In this case the test should be repeated in order to obtain new data.
	Data Security and Storage	Supplementary and raw data initially handwritten as text in lab books are not digitized but the information is transferred to digital reports stored in office computers. Experimental data concerning physicochemical analyses of water samples or characterization of materials (e.g. membranes) are stored in the hard drives of the corresponding instrument(s). Raw data collected during performance evaluation tests by SCADA systems of experimental devices are stored as .csv or spreadsheet (excel) files in lab computers. Reports, documents, photos, designs and other digitized material generated during – and relative to - intelWATT project are stored in office computers as well as in EmDESK platform. According to the lab policy all data are backed up regularly and the backups are stored in an internal network drive as well as removable hard drives. All the computers are protected by hard passwords, firewalls and antivirus / antimalware software. The current protocols prevent access to any external user.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 2– Experimental – WP2,3,4,9,10,11 – PARTNER: CNRS		
	Data utility	The data are useful for the whole CNRS team, to all members of the intelWATT project as well as to other research groups focusing in membrane science/technology and relevant applications. End users in the electric power generation and renewable energy fields, water and wastewater treatment professionals, process engineers and post graduate students focusing in the activities of intelWATT are indicative additional groups which benefit from the project's results. The data can be further exploited as source for new funding or new internal scientific projects as well as for commercialization purposes.
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Discoverability of the data is achieved by making the analyzed data available (provided that rules and provisions of signed GA, CA and IPR interests are not violated) for publication in scientific journals. Discoverability is further maximized linking the article with the relevant dataset (including related metadata) which is deposited in a scientific data repository. These data can be further exploited as source for new funding or new internal scientific projects. The relative metadata are stored mainly in .txt, and .pdf format. Experimental keywords will be used to create the corresponding metadata. These will include some common fields like the name/code of the material (membrane or IX resin), partner's acronym, as well as some specific info depending on the technique. For example, for membrane evaluation the following fields will be included: Type of experiment, Feed composition, Temperature, Pressure, Flow rates, Sample info, etc. Even though the metadata file is not created by a standard schema and is in the form of a simple .txt file, it should provide the following information: Title - Name of the dataset or research project that produced it Creator - Names and addresses of the organization or people who created the data Identifier - Number used to identify the data Subject - Keywords or phrases describing the subject or content of the data Access information - Where and how your data can be accessed by other researchers Date - Key dates associated with the data, including time period covered by the data Methodology - How the data was generated, including equipment or software used, experimental protocol, other things one might include in a lab notebook Data processing - Information on how the data has been altered or processed

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 2– Experimental – WP2,3,4,9,10,11 – PARTNER: CNRS		
		Sources - Citations to material for data derived from other sources, including details of where the source data is held and how it was accessed List of file names - List of all data files associated with the dataset, with their names and file extensions File Formats - Format(s) of the data, e.g. FITS, SPSS, HTML, JPEG, and any software required to read the data File structure - Organization of the data file(s) and the layout of the variables, when applicable Variable list - List and description of variables in the data files, when applicable Code lists - Explanation of codes or abbreviations used in either the file names or the variables in the data files (e.g. '999 indicates a missing value in the data')
	Identifiability of data (refer to standard id mechanisms)	Data is made identifiable by manually naming data files according to the rules of the lab and the file naming conventions referred below. No standard ID mechanism is used. Data (in the form of document, spreadsheets, images, presentation etc.) uploaded to an online data repository is identified by a Digital Object Identification (DOI) which assigned automatically by the repository. The DOI also contains metadata that provides the relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored.
	Naming conventions used	All the results and data generated during the implementation period of the project will be stored in dedicated folders. The folders' structure is tree like. The root folder (named with the project acronym "intelWATT") contains the WP folders. Further structuring inside the WP folders will be based on Task or Case study numbering. For example folder ..\\intelWATT\WP_2\TASK_3 contains files and data relative to the task 3 of WP 2. Data files are manually named in a way that their source can be uniquely distinguished. For example, a name consisting of a prefix (e.g. characteristic of the experimental technique or material) followed by a date and/or time stamp could be an appropriate convention for a file containing raw data from an experimental device's SCADA systems. In addition, the raw data files should be linked with relevant records (in physical form in a lab book or in a text file) containing all the necessary details for the reusability / interoperability of the data (e.g. the experimental set up and conditions, the preparation method of the materials, pre and / or post treatments of the samples etc.).
	Search keywords approach	In some cases some keywords are used including the name of the sample and the corresponding technique used.
	Clear versioning approach	Revised datasets will take the extension "rev" followed by the number of the revision made.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 2– Experimental – WP2,3,4,9,10,11 – PARTNER: CNRS		
	Standards or procedures for metadata creation applied	N/A
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	N/A
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	The raw data obtained by the CNRS team belongs to the CNRS. The research team will define the way and the time that these data can be presented. The analyzed data that are presented to the consortium are available for the whole consortium for the full duration of the project and according to the rules and provisions of the signed G.A., C.A. and intellectual property obligations. Accessibility of data out of the scope of intelWATT project requires the permission of CNRS. The same occurs also for the raw data. In such a case the raw data can be made available only upon request and under separate agreement between the partners and as long this request does not interfere with future publication and intellectual property interests, or precede data analysis.
	How data will be made available	Data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Disclosable data will be deposited in an open data repository, such as Zenodo or other recommended repositories (see below) provided that rules and provisions in signed GA and CA, IPR obligations or/and confidentiality restrictions are not violated. An option being considered is to link the data and metadata to forthcoming publications. Specific data can be made available upon requests (see 2.4)
	Methods or SW tools for data access	Tools for accessing the data are any text editor and spreadsheet packages for the output files. Optionally photo/image viewers / editors or CAD software may be needed for the visualization of some engineering designs.
	SW documentation and other information needed	In some cases, the supplementary notes may be needed for better understanding of the data.
	Repository for deposit of data, metadata, documentation and code	Specific repositories that will be considered include: FigShare, Mendeley Data, Open Science Framework, Harvard Dataverse, Dryad Digital Repository, Science Data Bank, Zenodo, which can be used for storage of data and metadata. These generalist repositories may also be appropriate for archiving associated analyses, or experimental-control data.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 2– Experimental – WP2,3,4,9,10,11 – PARTNER: CNRS		
	Access restrictions	As a rule, the availability of the data is according to the rule and the provisions of the signed G.A. and C.A., especially to the IPR clauses indicated there. In some circumstances, CNRS will not be able to disclose some of this data until a certain time within the project timeframe due to possible commercial confidentiality and intellectual property reasons.
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A
	Standard vocabulary or mapping to commonly used ontologies	N/A
	Information Portability and exporting	N/A
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)	**Data licensing for wide reuse**	N/A
	Timing of data availability for re-use (incl. indications on embargo)	Data can be made available after publication in scientific papers. The raw data will be not available to other partners, but can be shared only after permission of the CNRS team.
	Data usability by Third Parties (after the end of the project)	Except for the case where results have been published, data could be reused only with the permission of CNRS and according to the rules established in the G.A. and C.A.
	Restrictions to data re-use	Restrictions to the re-use of data from the Consortium Partners as imposed by the signed G.A. and C.A. rules, specific IPR established agreements and/or possible commercial confidentialities.
	Quality assurance process	Data quality is assured through the : - Regular checking and maintenance of the experimental and mechanical equipment including electronic sensors and control devices. - Calibration of sensors (temperature, pressure, flow, conductivity etc). - Calibration and performance verification of analytical instruments. - Use of standards for certain measurements. - Blank experiments and repetition of experiments. - Data entry validation and peer review of data.
	Length of time of data re-usability	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 2– Experimental – WP2,3,4,9,10,11 – PARTNER: CNRS		
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	N/A
	Data Management Responsibilities	The CNRS team is responsible for the management of the data obtained within the frame of intelWATT project. Researchers of CNRS are highly educated and qualified to ensure that dataset is accurate and reliable.
	Long Term Preservation assessment	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

4.1.3 DATA SET n. 3 – Experimental – WP 2,3,4,6 – PARTNER: CNR

DATA SET n. 3 – Experimental– WP 2,3,4,6 – PARTNER: CNR		
1 DATA SUMMARY	**Purpose of the Data**	Data collected relate to the effect of the physicochemical characteristics of water streams (dilute and concentrate) for the case study No2 (as indicated and supplied by Partner No15, ACSA, Spain) on membranes and membrane operations performance; lab scale development of the integrated RED/MD treatment process; design and performance evaluation studies (relative to process conditions, technologies combinations, modules configurations and setups, membrane shape and dimension etc.) at the lab and pilot scale. All the generated data are in close relation with the implementation of the project's deliverables and objectives.
	Type and Format of data	▪ Text: field or laboratory notes, survey responses – in plain text, (txt), HTML, XLM, PDF/A (performance evaluation tests and membrane characterization), ▪ Numeric: tables, counts, measurements – in .XLSX, .CSV (performance evaluation tests), ▪ Audiovisual: images, video – in .JPEG, .JPG, .PNG, .TIFF, AIFF, WAVE, .MP3, .MP4 (microscopic techniques SEM, AFM) ▪ Instrument specific: equipment outputs (physicochemical and electrochemical characterization and analysis) can be converted in XLS, ASCII and CSV format,
	Reused-Data	The majority of the data will be new, generated during the implementation of the project.
	Data origin	▪ Observational – data captured in real time (performance evaluation tests, membrane characterization), usually irreplaceable - i.e. sensor readings, images, sample data. The data are reproducible if the same conditions and equipment are used. ▪ Experimental – data from lab equipment (physicochemical analysis, membrane characterization), often reproducible - i.e. chromatograms, absorbance spectra, impedance spectra, porometer, SEM images. The observational and the experimental data are subject to analysis. In some case the analysis is very time consuming.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 3 – Experimental– WP 2,3,4,6 – PARTNER: CNR		
	Data size	Physicochemical analyses: - Revisable - Data file size depending on the analytical technique and instrument. Estimated file size will vary from some Kbytes to few Mbytes. Membrane characterization: - Revisable, - Collection of data files from various laboratory instruments like SEM, EIS, contact angle, etc. Data files size depends on the instrument and (regarding microscopies) from the resolution of the images. Estimated file sizes from some KB to few MB (images) per measurement. Performance evaluation tests: - Revisable, - Files of raw data acquired by the SCADA systems of experimental devises, controllers, online sensors and analyzers etc. ASCII files (in csv format) of approximately 10 MB per study depending on testing duration and sampling rate with an overall estimated size 50 GB , - The performance evaluation provides data that can be amended. In this case the test should be repeated in order to obtain new data.
	Data Security and Storage	Supplementary and raw data initially handwritten as text in lab books are not digitized but the information is transferred to digital reports stored in office computers. Experimental data concerning physicochemical analyses of water samples or characterization of membranes are stored in the hard drives of the corresponding instrument(s). Raw data collected during performance evaluation tests by SCADA systems of experimental devices are stored as .csv or spreadsheet (excel) files in lab computers. Reports, documents, photos, designs and other digitized material generated during – and relative to - intelWATT project are stored in office computers as well as in EmDESK platform. All data are backed up regularly and the backups are stored in an internal cloud storage as well as removable hard drives. All the computers are protected by passwords, firewalls and antivirus / antimalware software. The current protocols prevent access to any external user.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 3 – Experimental– WP 2,3,4,6 – PARTNER: CNR		
	Data utility	The data are useful for the whole team of CNR, to all members of the intelWATT project as well as to other research groups focusing in membrane science/technology and relevant applications. End users in the electric power generation and renewable energy fields, water and wastewater treatment professionals, process engineers and post graduate students focusing in the activities of intelWATT are indicative additional groups which benefit from the project's results. The data can be further exploited as source for new funding or new internal scientific projects as well as for commercialization purposes.
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Discoverability of the data is achieved by making the analyzed data available (provided that rules and provisions of signed GA, CA and IPR interests are not violated) for publication in scientific journals. Discoverability is further maximized linking the article with the relevant dataset (including related metadata) which is deposited in a scientific data repository. These data can be further exploited as source for new fundings or new internal scientific projects. The relative metadata are stored mainly in .txt, and .pdf format. Experimental keywords will be used to create the corresponding metadata. These will include some common fields like the name/code of the material (membrane), partner's acronym, as well as some specific info depending on the technique. For example, for membrane evaluation the following fields will be included: Type of experiment, Feed composition, Temperature, Pressure, Flow rates, Sample info, etc. Even though the metadata file is not created by a standard schema and is in the form of a simple .txt file, it should provide the following information: Title - Name of the dataset or research project that produced it Creator - Names and addresses of the organization or people who created the data Identifier - Number used to identify the data Subject - Keywords or phrases describing the subject or content of the data Access information - Where and how your data can be accessed by other researchers Date - Key dates associated with the data, including time period covered by the data Methodology - How the data was generated, including equipment or software used, experimental protocol, other things one might include in a lab notebook Data processing - Information on how the data has been altered or processed

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 3 – Experimental– WP 2,3,4,6 – PARTNER: CNR		
		Sources - Citations to material for data derived from other sources, including details of where the source data is held and how it was accessed. List of file names - List of all data files associated with the dataset, with their names and file extensions. File Formats - Format(s) of the data, e.g. FITS, SPSS, HTML, JPEG, and any software required to read the data. File structure - Organization of the data file(s) and the layout of the variables, when applicable. Variable list - List and description of variables in the data files, when applicable. Code lists - Explanation of codes or abbreviations used in either the file names or the variables in the data files (e.g. '"999 indicates a missing value in the data").
	Identifiability of data (refer to standard id mechanisms)	Data is made identifiable by manually naming data files according to the rules of the lab and the file naming conventions referred below. No standard ID mechanism is used. Data (in the form of document, spreadsheets, images, presentation etc.) uploaded to an online data repository is identified by a Digital Object Identification (DOI) which assigned automatically by the repository. The DOI also contains metadata that provides the relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored.
	Naming conventions used	All the results and data generated during the implementation period of the project will be stored in dedicated folders. The folders' structure is tree like. The root folder (named with the project acronym "intelWATT") contains the WP folders. Further structuring inside the WP folders will be based on Task or Case study numbering. For example folder ..\\intelWATT\WP_2\TASK_3 contains files and data relative to the task 3 of WP 2. Data files are manually named in a way that their source can be uniquely distinguished. In addition, the raw data files should be linked with relevant records (in physical form in a lab book or in a text file) containing all the necessary details for the reusability / interoperability of the data (e.g. the experimental set up and conditions, pre and / or post treatments of the samples etc.).
	Search keywords approach	In some cases some keywords are used including the name of the sample and the corresponding experiment.
	Clear versioning approach	Revised datasets will take the extension "rev" followed by the number of the revision made.
	Standards or procedures for metadata creation applied	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 3 – Experimental– WP 2,3,4,6 – PARTNER: CNR		
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	N/A
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	The raw data obtained by the CNR team belongs to the CNR. The research team will define the way and the time that these data can be presented. The analyzed data that are presented to the consortium are available for the whole consortium for the full duration of the project and according to the rules and provisions of the signed G.A., C.A. and intellectual property obligations. Accessibility of data out of the scope of intelWATT project requires the permission of CNR. The same occurs also for the raw data. In such a case the raw data can be made available only upon request and under separate agreement between the partners and as long this request does not interfere with future publication and intellectual property interests, or precede data analysis.
	How data will be made available	Data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Disclosable data will be deposited in an open data repository, such as Zenodo or other recommended repositories (see below) provided that rules and provisions in signed GA and CA, IPR obligations or/and confidentiality restrictions are not violated. An option being considered is to link the data and metadata to forthcoming publications. Specific data can be made available upon requests (see 2.4)
	Methods or SW tools for data access	Tools for accessing the data are any text editor and spreadsheet packages for the output files. Optionally photo/image viewers / editors or CAD software may be needed for the visualization of some engineering designs.
	SW documentation and other information needed	In some cases, the supplementary notes may be needed for better understanding of the data.
	Repository for deposit of data, metadata, documentation and code	Specific repositories that will be considered include: Zenodo, FigShare, Mendeley Data, Open Science Framework, Harvard Dataverse, Dryad Digital Repository, Science Data Bank, which can be used for storage of data and metadata. These generalist repositories may also be appropriate for archiving associated analyses, or experimental-control data.
	Access restrictions	As a rule, the availability of the data is according to the rule and the provisions of the signed G.A. and C.A., especially to the IPR clauses indicated there. In some circumstances, CNR will not be able to disclose some of this data until a certain time within the project timeframe due to possible commercial confidentiality and intellectual property reasons.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 3 – Experimental– WP 2,3,4,6 – PARTNER: CNR		
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A
	Standard vocabulary or mapping to commonly used ontologies	N/A
	Information Portability and exporting	N/A
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)	**Data licensing for wide reuse**	N/A
	Timing of data availability for re-use (incl. indications on embargo)	Data can be made available after publication in scientific papers. The raw data will be not available to other partners, but can be shared only after permission of the CNR team.
	Data usability by Third Parties (after the end of the project)	Except for the case where results have been published, data could be reused only with the permission of CNR and according to the rules established in the G.A. and C.A.
	Restrictions to data re-use	Restrictions to the re-use of data from the Consortium Partners as imposed by the signed G.A. and C.A. rules, specific IPR established agreements and/or possible commercial confidentialities.
	Quality assurance process	Data quality is assured through the : - Regular checking and maintenance of the experimental and mechanical equipment including electronic sensors and control devices. - Calibration of sensors (temperature, pressure, flow, conductivity etc.) - Calibration and performance verification of analytical instruments - Use of standards for certain measurements - Blank experiments and repetition of experiments. - Data entry validation and peer review of data.
	Length of time of data re-usability	N/A
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	N/A
	Data Management Responsibilities	The CNR team is responsible for the management of the data obtained within the frame of intelWATT project. Researchers of CNR are highly educated and qualified to ensure that dataset is accurate and reliable.
	Long Term Preservation assessment	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

4.1.4 DATA SET n. 4 – Experimental – WP2,5 – PARTNER: PPC

DATA SET n. 4 – Experimental– WP2,5 – PARTNER: PPC		
1 DATA SUMMARY	**Purpose of the Data**	Data concerns physicochemical analyses of relevant water streams for the case study No1, indicated and supplied by PPC to the partners. Moreover, during pilot case study 1, data will be produced regarding the performance evaluation of the pilot materials relative to process conditions (especially for fouling and corrosion), technologies combinations, modules configurations and setups, membrane shape and dimension etc. at the lab but mostly at pilot scale. Furthermore process design drawing(s) will be derived. All the generated data are in close relation with the implementation of the project's deliverables and objectives.
	Type and Format of data	▪ Text: use case field or laboratory notes, survey responses – in plain text, (txt), HTML, XLM, PDF/A {performance evaluation tests and material preparation}, ▪ Numeric: tables, counts, measurements – in .XLSX, .CSV (performance evaluation tests), ▪ Audiovisual: images, video – in .JPEG, .JPG, .PNG, .TIFF, AIFF, WAVE, .MP3, .MP4 (microscopic techniques SEM, AFM), ▪ Instrument specific: equipment outputs (specify equipment and format) {physicochemical characterization and analysis - can be converted in XLS, ASCII and CSV format}.
	Reused-Data	The majority of the data will be new, generated during the implementation of the project.
	Data origin	▪ Observational – data captured in real time (performance evaluation tests, material preparation), usually irreplaceable - i.e. sensor readings, images, telemetries, sample data. The data are reproducible if the same conditions and equipment are used. ▪ Experimental – data from the use case field generated through lab equipment (physicochemical analysis, materials characterization), often reproducible - i.e. chromatograms, absorbance spectra, x-ray diffraction data, porometer, porosimeter. The observational and the experimental data are subject to analysis. In some case the analysis is very time consuming which increases the cost.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 4 – Experimental– WP2,5 – PARTNER: PPC		
	Data size	Physicochemical analyses: - Revisable - Data file size depending on the analytical technique and instrument. The Datasets vary from some Kbytes (AAS) to few Mbytes (Ion chromatography) per experiment. Performance evaluation tests: - Revisable - Files of raw data acquired by the SCADA systems of experimental devises, controllers, online sensors and analyzers etc. ASCII files (in csv format) of approximately 10 MB per study depending on testing duration and sampling rate with an overall estimated size 50 GB . - The performance evaluation provides data that can be amended. In this case the test should be repeated in order to obtain new data.
	Data Security and Storage	Supplementary and raw data initially handwritten as text are not digitized but the information is transferred to digital reports stored in office computers. Experimental data concerning physicochemical analyses of water samples or characterization of materials (e.g. membranes) are stored in the hard drives of the corresponding instrument(s). Raw data collected during performance evaluation tests by SCADA systems of experimental devices are stored as .csv or spreadsheet (excel) files in lab computers. Reports, documents, photos, designs and other digitized material generated during – and relative to - intelWATT project are stored in office computers as well as in EmDESK platform. According to the company policy all data are backed up regularly and the backups are stored in an internal network drive as well as removable hard drives. All the computers are protected by hard passwords, firewalls and antivirus / antimalware software. The current protocols prevent access to any external user.
	Data utility	The data are useful for the whole team of PPC and to all members of the intelWATT project participating to the use case 1 apllications. Water and wastewater treatment users, process engineers and post graduate students focusing in the activities of intelWATT are indicative additional groups which benefit from the project's results. The data can be further exploited as source for new funding or new internal scientific projects.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 4 – Experimental– WP2,5 – PARTNER: PPC		
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Discoverability of the data is achieved by making the analyzed data available (provided that rules and provisions of signed GA, CA and IPR interests are not violated) for publication in scientific journals. Discoverability is further maximized linking the article with the relevant dataset (including related metadata) which is deposited in a scientific data repository. These data can be further exploited as source for new funding or new internal scientific projects. The relative metadata are stored mainly in .txt, and .pdf format. Experimental keywords will be used to create the corresponding metadata. These will include some common fields like the name/code of the material installed or water stream treated, partner's acronym, as well as some specific info depending on the technique. For example, for membrane evaluation the following fields will be included: Type of experiment, Feed composition, Temperature, Pressure, Flow rates, Sample info, etc. Even though the metadata file is not created by a standard schema and is in the form of a simple .txt file, it should provide the following information: Title - Name of the dataset or research project that produced it. Creator - Names and addresses of the organization or people who created the data Identifier - Number used to identify the data. Subject - Keywords or phrases describing the subject or content of the data Access information - Where and how your data can be accessed by other researchers Date - Key dates associated with the data, including time period covered by the data Methodology - How the data was generated, including equipment or software used, experimental protocol, other things one might include in a lab notebook. Data processing - Information on how the data has been altered or processed

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 4 – Experimental– WP2,5 – PARTNER: PPC		
		Sources - Citations to material for data derived from other sources, including details of where the source data is held and how it was accessed. List of file names - List of all data files associated with the dataset, with their names and file extensions. File Formats - Format(s) of the data, e.g. FITS, SPSS, HTML, JPEG, and any software required to read the data. File structure - Organization of the data file(s) and the layout of the variables, when applicable. Variable list - List and description of variables in the data files, when applicable. Code lists - Explanation of codes or abbreviations used in either the file names or the variables in the data files (e.g. '999 indicates a missing value in the data').
	Identifiability of data (refer to standard id mechanisms)	Data is made identifiable by manually naming data files according to the rules of the lab and the file naming conventions referred below. No standard ID mechanism is used. Data (in the form of document, spreadsheets, images, presentation etc.) uploaded to an online data repository is identified by a Digital Object Identification (DOI) which assigned automatically by the repository. The DOI also contains metadata that provides the relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored.
	Naming conventions used	Data are manually named in a way that one can distinguish their source. Also the name of the interpreter/analyser is added. These data are used for the preparation of internal reports. The reports are named by the project and the period, in which the data belongs.
	Search keywords approach	In some cases some keywords are used including the name of the water stream sample and the corresponding technique used.
	Clear versioning approach	The versioning of data files is manual indicating the date of its creation and the writer.
	Standards or procedures for metadata creation applied	N/A
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 4 – Experimental– WP2,5 – PARTNER: PPC		
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	The raw data obtained and derived by the PPC's processes belongs to the PPC. The research team will define the way and the time that these data can be presented. The analyzed data that are presented to the consortium are available for the whole consortium for the full duration of the project and according to the rules and provisions of the signed G.A., C.A. and intellectual property obligations. Accessibility of data out of the scope of intelWATT project requires the permission of PPC. The same occurs also for the raw data. In such a case the raw data can be made available only upon request and under separate agreement between the partners and as long this request does not interfere with future publication and intellectual property interests or precede data analysis.
	How data will be made available	Data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Specific data can be made available upon requests (see 2.4)
	Methods or SW tools for data access	Tools for accessing the data are any text editor and spreadsheet packages for the output files. Optionally photo/image viewers / editors or CAD software may be needed for the visualization of some engineering designs.
	SW documentation and other information needed	In some cases, the supplementary notes may be needed for better understanding of the data.
	Repository for deposit of data, metadata, documentation and code	N/A
	Access restrictions	As a rule, the availability of the data is according to the rule and the provisions of the signed G.A. and C.A., especially to the IPR clauses indicated there. In some circumstances, PPC will not be able to disclose some of this data until a certain time within the project timeframe due to possible commercial confidentiality and intellectual property reasons.
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A
	Standard vocabulary or mapping to commonly used ontologies	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 4 – Experimental– WP2,5 – PARTNER: PPC		
	Information Portability and exporting	N/A
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)	**Data licensing for wide reuse**	N/A
	Timing of data availability for re-use (incl. indications on embargo)	Data can be made available after publication in scientific papers. The raw data will be not available to other partners, but can be shared only after permission of the PPC team.
	Data usability by Third Parties (after the end of the project)	Except for the case where results have been published, data could be reused only with the permission of PPC and according to the rules established in the G.A. and C.A.
	Restrictions to data re-use	Restrictions to the re-use of data from the Consortium Partners as imposed by the signed G.A. and C.A. rules, specific IPR established agreements and/or possible commercial confidentialities.
	Quality assurance process	Data quality is assured through the: - Regular checking and maintenance of the experimental and mechanical equipment including electronic sensors and control devices, - Calibration of sensors (temperature, pressure, flow, conductivity etc.). - Calibration and performance verification of analytical instruments, - Use of standards for certain measurements, - Blank experiments and repetition of experiments, - Data entry validation and peer review of data.
	Length of time of data re-usability	N/A
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	N/A
	Data Management Responsibilities	The PPC team is responsible for the management of the data obtained from PPC processes within the frame of intelWATT project.
	Long Term Preservation assessment	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

4.1.5 DATA SET n. 5 – Experimental – WP2,3,4,7 – PARTNER: THK

DATA SET n. 5 – Experimental– WP2,3,4,7 – PARTNER: THK		
1 DATA SUMMARY	**Purpose of the Data**	Data concerns physicochemical analyses of relevant water streams for the case study No 3 (as indicated and supplied by Partner No 10, BIA Kunststoff- und Galvanotechnik GmbH & Co. KG, material characterization (such as ion exchange resins or activated carbon), membrane module design development and performance evaluation studies (relative to process conditions, technologies combinations, modules configurations and setups, membrane shape and dimension etc.) at lab and pilot scale, process design drawing(s). All the generated data are in close relation with the implementation of the project's deliverables and objectives.
	Type and Format of data	▪ Text: field or laboratory notes, survey responses – in plain text, (txt), HTML, XLM, PDF/A {performance evaluation tests and material preparation}, ▪ Numeric: tables, counts, measurements – in .XLSX, .CSV (performance evaluation tests), ▪ Audiovisual: images, video – in .JPEG, .JPG, .PNG, .TIFF, AIFF, WAVE, .MP3, .MP4 (microscopic techniques SEM, AFM), ▪ Instrument specific: equipment outputs (specify equipment and format) {physicochemical characterization and analysis - can be converted in XLS, ASCII and CSV format}.
	Reused-Data	The majority of the data will be new, generated during the implementation of the project.
	Data origin	▪ Observational – data captured in real time (performance evaluation tests, material preparation), usually irreplaceable - i.e. sensor readings, images, telemetries, sample data. The data are reproducible if the same conditions and equipment are used. ▪ Experimental – data from lab equipment (physicochemical analysis, materials characterization), often reproducible - i.e. chromatograms, absorbance spectra, x-ray diffraction data, porometer, porosimeter. The observational and the experimental data are subject to analysis. In some cases, the analysis is very time consuming which increases the cost.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 5 – Experimental– WP2,3,4,7 – PARTNER: THK		
	Data size	Physicochemical analyses: - Revisable, - Data file size depending on the analytical technique and instrument. The Datasets vary from some Kbytes (AAS) to few Mbytes (Ion chromatography) per experiment. Material characterization: - Revisable, - Collection of data files from various laboratory instruments like SEM, AFM, contact angle, surface tension etc. Data files size depends on the instrument and (regarding microscopies) from the resolution of the images. Estimated file sizes from some KB to few MB (images) per measurement. Performance evaluation tests: - Revisable, - Files of raw data acquired by the SCADA systems of experimental devises, controllers, online sensors and analyzers etc. ASCII files (in csv format) of approximately 10 MB per study depending on testing duration and sampling rate with an overall estimated size 50 GB. - The performance evaluation provides data that can be amended. In this case the test should be repeated in order to obtain new data.
	Data Security and Storage	Supplementary and raw data initially handwritten as text in lab books are not digitized but the information is transferred to digital reports stored in office computers. Experimental data concerning physicochemical analyses of water samples or characterization of materials (e.g. membranes) are stored in the hard drives of the corresponding instrument(s). Raw data collected during performance evaluation tests by SCADA systems of experimental devices are stored as .csv or spreadsheet (excel) files in lab computers. Reports, documents, photos, designs and other digitized material generated during – and relative to - intelWATT project are stored in office computers as well as in EmDESK platform. According to the lab policy all data are backed up regularly and the backups are stored in an internal network drive as well as removable hard drives. All the computers are protected by hard passwords, firewalls and antivirus / antimalware software. The current protocols prevent access to any external user.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 5 – Experimental– WP2,3,4,7 – PARTNER: THK		
	Data utility	The data are useful for the whole team of THK, to all members of the intelWATT project as well as to other research groups focusing in membrane science/technology and relevant applications. End users in the electric power generation and renewable energy fields, water and wastewater treatment professionals, process engineers and post graduate students focusing in the activities of intelWATT are indicative additional groups which benefit from the project's results. The data can be further exploited as source for new funding or new internal scientific projects as well as for commercialization purposes.
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Discoverability of the data is achieved by making the analyzed data available (provided that rules and provisions of signed GA, CA and IPR interests are not violated) for publication in scientific journals. Discoverability is further maximized linking the article with the relevant dataset (including related metadata) which is deposited in a scientific data repository. These data can be further exploited as source for new funding or new internal scientific projects. The relative metadata are stored mainly in .txt, and .pdf format. Experimental keywords will be used to create the corresponding metadata. These will include some common fields like the name/code of the material (membrane or IX resin), partner's acronym, as well as some specific info depending on the technique. For example, for membrane evaluation the following fields will be included: Type of experiment, Feed composition, Temperature, Pressure, Flow rates, Sample info, etc. Even though the metadata file is not created by a standard schema and is in the form of a simple .txt file, it should provide the following information: Title - Name of the dataset or research project that produced it. Creator - Names and addresses of the organization or people who created the data. Identifier - Number used to identify the data. Subject - Keywords or phrases describing the subject or content of the data. Access information - Where and how your data can be accessed by other researchers. Date - Key dates associated with the data, including time period covered by the data. Methodology - How the data was generated, including equipment or software used, experimental protocol, other things one might include in a lab notebook. Data processing - Information on how the data has been altered or processed.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 5 – Experimental– WP2,3,4,7 – PARTNER: THK		
		Sources - Citations to material for data derived from other sources, including details of where the source data is held and how it was accessed. List of file names - List of all data files associated with the dataset, with their names and file extensions. File Formats - Format(s) of the data, e.g. FITS, SPSS, HTML, JPEG, and any software required to read the data. File structure - Organization of the data file(s) and the layout of the variables, when applicable. Variable list - List and description of variables in the data files, when applicable. Code lists - Explanation of codes or abbreviations used in either the file names or the variables in the data files (e.g. '999 indicates a missing value in the data').
	Identifiability of data (refer to standard id mechanisms)	Data is made identifiable by manually naming data files according to the rules of the lab and the file naming conventions referred below. No standard ID mechanism is used. Data (in the form of document, spreadsheets, images, presentation etc.) uploaded to an online data repository is identified by a Digital Object Identification (DOI) which assigned automatically by the repository. The DOI also contains metadata that provides the relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored.
	Naming conventions used	All the results and data generated during the implementation period of the project will be stored in dedicated folders. The folders' structure is tree like. The root folder (named with the project acronym "intelWATT") contains the WP folders. Further structuring inside the WP folders will be based on Task or Case study numbering. For example folder ..\\intelWATT\WP_2\TASK_3 contains files and data relative to the task 3 of WP 2. Data files are manually named in a way that their source can be uniquely distinguished. For example, a name consisting of a prefix (e.g. characteristic of the experimental technique or material) followed by a date and/or time stamp could be an appropriate convention for a file containing raw data from an experimental device's SCADA systems. In addition, the raw data files should be linked with relevant records (in physical form in a lab book or in a text file) containing all the necessary details for the reusability / interoperability of the data (e.g. the experimental set up and conditions, the preparation method of the materials, pre and / or post treatments of the samples etc.).
	Search keywords approach	In some cases some keywords are used including the name of the sample and the corresponding technique used.
	Clear versioning approach	Revised datasets will take the extension "rev" followed by the number of the revision made.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 5 – Experimental– WP2,3,4,7 – PARTNER: THK		
	Standards or procedures for metadata creation applied	N/A
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	N/A
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	The raw data obtained by the THK team belongs to THK. The research team will define the way and the time that these data can be presented. The analyzed data that are presented to the consortium are available for the whole consortium for the full duration of the project and according to the rules and provisions of the signed G.A., C.A. and intellectual property obligations. Accessibility of data out of the scope of intelWATT project requires the permission of THK. The same occurs also for the raw data. In such a case the raw data can be made available only upon request and under separate agreement between the partners and as long this request does not interfere with future publication and intellectual property interests, or precede data analysis.
	How data will be made available	Data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Disclosable data will be deposited in an open data repository, such as Zenodo or other recommended repositories (see below) provided that rules and provisions in signed GA and CA, IPR obligations or/and confidentiality restrictions are not violated. An option being considered is to link the data and metadata to forthcoming publications. Specific data can be made available upon requests (see 2.4)
	Methods or SW tools for data access	Tools for accessing the data are any text editor and spreadsheet packages for the output files. Optionally photo/image viewers / editors or CAD software may be necessary for the visualization of some engineering designs.
	SW documentation and other information needed	In some cases, the supplementary notes may be necessary for better understanding of the data.
	Repository for deposit of data, metadata, documentation and code	Specific repositories that will be considered include: FigShare, Mendeley Data, Open Science Framework, Harvard Dataverse, Dryad Digital Repository, Science Data Bank, Zenodo, which can be used for storage of data and metadata. These generalist repositories may also be appropriate for archiving associated analyses, or experimental-control data.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 5 – Experimental– WP2,3,4,7 – PARTNER: THK		
	Access restrictions	As a rule, the availability of the data is according to the rule and the provisions of the signed G.A. and C.A., especially to the IPR clauses indicated there. In some circumstances, THK will not be able to disclose some of this data until a certain time within the project timeframe due to possible commercial confidentiality and intellectual property reasons.
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A
	Standard vocabulary or mapping to commonly used ontologies	N/A
	Information Portability and exporting	N/A
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)	**Data licensing for wide reuse**	N/A
	Timing of data availability for re-use (incl. indications on embargo)	Data can be made available after publication in scientific papers. The raw data will be not available to other partners, but can be shared only after permission of the THK team.
	Data usability by Third Parties (after the end of the project)	Except for the case where results have been published, data could be reused only with the permission of THK and according to the rules established in the G.A. and C.A.
	Restrictions to data re-use	Restrictions to the re-use of data from the Consortium Partners as imposed by the signed G.A. and C.A. rules, specific IPR established agreements and/or possible commercial confidentialities
	Quality assurance process	Data quality is assured through the : - Regular checking and maintenance of the experimental and mechanical equipment including electronic sensors and control devices, - Calibration of sensors (temperature, pressure, flow, conductivity etc.), - Calibration and performance verification of analytical instruments, - Use of standards for certain measurements, - Blank experiments and repetition of experiments, - Data entry validation and peer review of data.
	Length of time of data re-usability	N/A
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 5 – Experimental– WP2,3,4,7 – PARTNER: THK		
	Data Management Responsibilities	The THK team is responsible for the management of the data obtained within the frame of intelWATT project. Researchers of THK are high educated and qualified to ensure that dataset is accurate and reliable.
	Long Term Preservation assessment	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

4.1.6 DATA SET n. 6 – Experimental – WP3 – PARTNER: UOB

DATA SET n. 6 – Experimental – WP3 – PARTNER: UOB		
1 DATA SUMMARY	**Purpose of the Data**	Data concerns real time performance data logged from experimental desalination rig in the lab. This data is used to assess the performance and durability of the separation process, and how reliable it will be in the target application. All the generated data are in close relation with the implementation of the project's deliverables and objectives in WP3.
	Type and Format of data	Numeric: tables, counts, measurements – in .XLSX, .CSV. These typically include sensor outputs of parameters such as pressure, conductivity, temperature, flow rate etc that are important for the understanding and control of the process, related mainly to case study 3. Models, computational models, computer code - (specify type and format) will be in Matlab or Excel format, are used to represent and predict the performance of the desalination equipment after verification against the measured values.
	Reused-Data	N/A
	Data origin	Observational and experimental– data captured in real time (performance evaluation tests), usually irreplaceable - i.e. sensor readings, and manual observations. The data are reproducible if the same conditions and equipment are used. The observational and the experimental data are subject to analysis. In some case the analysis is very time consuming which increases the cost.
	Data size	Performance evaluation tests: - Revisable, - Files of raw data acquired by the data logger systems of experimental devises, controllers, online sensors and analyzers etc. ASCII files (in csv format) of approximately 1 MB per study depending on testing duration and sampling rate with an overall estimated size 1 GB, - The performance evaluation provides data that can be amended. In this case the test should be repeated in order to obtain new data.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 6 – Experimental – WP3 – PARTNER: UOB		
	Data Security and Storage	Supplementary and raw data initially handwritten as text in lab books are not digitized but the information is transferred to digital reports stored in office computers. Experimental data concerning from desalination rig stored in the hard drives of the corresponding dedicated PC and backed up on the University server. Reports, documents, photos, designs and other digitized material generated during – and relative to - intelWATT project are stored in office computers as well as in EmDESK platform. According to the lab policy all data are backed up regularly and the backups are stored in an internal network drive and/or university-IT service via Birmingham Environment for Academic Research (BEAR) or approved cloud service. All the computers are protected by hard passwords, firewalls and antivirus / antimalware software. The current protocols prevent access to any external user.
	Data utility	The data are useful for the whole team of UoB, to all members of the intelWATT project as well as to other research groups focusing in membrane science/technology and relevant applications. End users in the electric power generation and renewable energy fields, water and wastewater treatment professionals, process engineers and post graduate students focusing in the activities of intelWATT are indicative additional groups which benefit from the project's results. The data can be further exploited as source for new funding or new internal scientific projects as well as for commercialization purposes.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 6 – Experimental – WP3 – PARTNER: UOB		
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Discoverability of the data is achieved by making the analyzed data available (provided that rules and provisions of signed GA, CA and IPR interests are not violated) for publication in scientific journals. Discoverability is further maximized linking the article with the relevant dataset (including related metadata) which is deposited in a scientific data repository. These data can be further exploited as source for new funding or new internal scientific projects. The relative metadata are stored mainly in .txt, and .pdf format. Experimental keywords will be used to create the corresponding metadata. These will include some common fields like the name/code of the material (membrane or IX resin), partner's acronym, as well as some specific info depending on the technique. For example, for membrane evaluation the following fields will be included: Type of experiment, Feed composition, Temperature, Pressure, Flow rates, Sample info, etc. Even though the metadata file is not created by a standard schema and is in the form of a simple .txt file, it should provide the following information: Title - Name of the dataset or research project that produced it. Creator - Names and addresses of the organization or people who created the data Identifier - Number used to identify the data. Subject - Keywords or phrases describing the subject or content of the data Access information - Where and how your data can be accessed by other researchers Date - Key dates associated with the data, including time period covered by the data Methodology - How the data was generated, including equipment or software used, experimental protocol, other things one might include in a lab notebook Data processing - Information on how the data has been altered or processed
	Identifiability of data (refer to standard id mechanisms)	Data is made identifiable by manually naming data files according to the rules of the lab and the file naming conventions referred below. No standard ID mechanism is used. Data (in the form of document, spreadsheets, images, presentation etc.) uploaded to an online data repository is identified by a Digital Object Identification (DOI) which assigned automatically by the repository. The DOI also contains metadata that provides the relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 6 – Experimental – WP3 – PARTNER: UOB		
	Naming conventions used	All the results and data generated during the implementation period of the project will be stored in dedicated folders. The folders' structure is tree like. The root folder (named with the project acronym "intelWATT") contains the WP folders. Further structuring inside the WP folders will be based on Task or Case study numbering. For example folder ..\\intelWATT\WP_3\TASK_3 contains files and data relative to the task 3 of WP 2. Data files are manually named in a way that their source can be uniquely distinguished. For example, a name consisting of a prefix (e.g. characteristic of the experimental technique or material) followed by a date and/or time stamp could be an appropriate convention for a file containing raw data from an experimental device's data-logging systems. In addition, the raw data files should be linked with relevant records (in physical form in a lab book or in a text file) containing all the necessary details for the reusability / interoperability of the data (e.g. the experimental set up and conditions, the preparation method of the materials, pre and / or post treatments of the samples etc.).
	Search keywords approach	No specific approach is provided. The pattern for keywords generation is mainly based on the kind of experiments performed, the experimental procedures and conditions, characteristics and properties of the under testing materials, analytical methods etc. As an example, for the performance evaluation of a PVDF membrane the recommended keywords are: PVDF membranes, flat sheet modules, membrane distillation, hydrophobicity, heat recovery etc.
	Clear versioning approach	The versioning of data files is manual indicating the date of its creation and the writer.
	Standards or procedures for metadata creation applied	N/A
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 6 – Experimental – WP3 – PARTNER: UOB		
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	The raw data obtained by the UoB team belongs to the UoB. The research team will define the way and the time that these data can be presented. The analyzed data that are presented to the consortium are available for the whole consortium for the full duration of the project and according to the rules and provisions of the signed G.A., C.A. and intellectual property obligations. Accessibility of data out of the scope of intelWATT project requires the permission of UoB. The same occurs also for the raw data. In such a case the raw data can be made available only upon request and under separate agreement between the partners and as long this request does not interfere with future publication and intellectual property interests, or precede data analysis.
	How data will be made available	Data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Disclosable data will be deposited in an open data repository, such as Zenodo or other recommended repositories (see below) provided that rules and provisions in signed GA and CA, IPR obligations or/and confidentiality restrictions are not violated. An option being considered is to link the data and metadata to forthcoming publications. Specific data can be made available upon requests (see 2.4)
	Methods or SW tools for data access	Tools for accessing the data are any text editor and spreadsheet packages for the output files. Optionally photo/image viewers / editors or CAD software may be needed for the visualization of some engineering designs.
	SW documentation and other information needed	In some cases, the supplementary notes may be needed for better understanding of the data.
	Repository for deposit of data, metadata, documentation and code	Specific repositories that will be considered include: FigShare, Mendeley Data, Open Science Framework, Harvard Dataverse, Dryad Digital Repository, Science Data Bank, Zenodo, which can be used for storage of data and metadata. These generalist repositories may also be appropriate for archiving associated analyses, or experimental-control data.
	Access restrictions	As a rule, the availability of the data is according to the rule and the provisions of the signed G.A. and C.A., especially to the IPR clauses indicated there. In some circumstances, UoB will not be able to disclose some of this data until a certain time within the project timeframe due to possible commercial confidentiality and intellectual property reasons.
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 6 – Experimental – WP3 – PARTNER: UOB		
	Standard vocabulary or mapping to commonly used ontologies	N/A
	Information Portability and exporting	Indicate whether documents should be exported massively in an external file.
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)	**Data licensing for wide reuse**	N/A
	Timing of data availability for re-use (incl. indications on embargo)	Data can be made available after publication in scientific papers. The raw data will be not available to other partners, but can be shared only after permis.sion of the UoB team.
	Data usability by Third Parties (after the end of the project)	Except for the case where results have been published, data could be reused only with the permission of UoB and according to the rules established in the G.A. and C.A.
	Restrictions to data re-use	Restrictions to the re-use of data from the Consortium Partners as imposed by the signed G.A. and C.A. rules, specific IPR established agreements and/or possible commercial confidentialities.
	Quality assurance process	Data quality is assured through the: - Regular checking and maintenance of the experimental and mechanical equipment including electronic sensors and control devices, - Calibration of sensors (temperature, pressure, flow, conductivity etc.), - Calibration and performance verification of analytical instruments, - Use of standards for certain measurements, - Blank experiments and repetition of experiments, - Data entry validation and peer review of data.
	Length of time of data re-usability	N/A
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	N/A
	Data Management Responsibilities	The UoB team is responsible for the management of the data obtained within the frame of intelWATT project. Researchers of UoB are high educated and qualified to ensure that dataset is accurate and reliable.
	Long Term Preservation assessment	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

4.1.7 DATA SET n. 7 – Experimental – WP2,3,4,8 – PARTNER: POLITO

DATA SET n. 7 – Experimental – WP2,3,4,8 – PARTNER: POLITO		
1 DATA SUMMARY	**Purpose of the Data**	Data concerns physicochemical characterization of graphene-based membrane, optimization of membrane preparation parameters for RED, membrane scaling up at the lab and pilot scale, sensors measurements in the different cases of study. All the generated data are in close relation with the implementation of the project's deliverables and objectives.
	Type and Format of data	▪ Text: field or laboratory notes, survey responses – in plain text, (txt), HTML, XLM, PDF/A {performance evaluation tests and material preparation }, ▪ Numeric: tables, counts, measurements – in .XLSX, .CSV (performance evaluation tests), ▪ Audiovisual: images, video – in .JPEG, .JPG, .PNG, .TIFF, AIFF, WAVE, .MP3, .MP4 (microscopic techniques SEM), ▪ Instrument specific: equipment outputs (specify equipment and format) {physicochemical characterization and analysis - can be converted in XLS, ASCII and CSV format}.
	Reused-Data	The majority of the data will be new, generated during the implementation of the project.
	Data origin	▪ Observational – data captured in real time (performance evaluation tests, material preparation), usually irreplaceable - i.e. sensor readings, images, telemetries, sample data. The data are reproducible if the same conditions and equipment are used. ▪ Experimental – data from lab equipment (physicochemical analysis, materials characterization), often reproducible - i.e. absorbance spectra, x-ray diffraction data, compositional data from XPS,etc. The observational and the experimental data are subject to analysis. In some case the analysis is very time consuming which increases the cost.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 7 – Experimental – WP2,3,4,8 – PARTNER: POLITO		
	Data size	Physicochemical analyses: - Revisable. - Data file size depending on the analytical technique and instrument. The Datasets vary from some Kbytes to few Mbytes per experiment. Material characterization: - Revisable. - Collection of data files from various laboratory instruments like SEM, XPS, Infrared spectrophotometer, contact angle, surface tension etc. Data files size depends on the instrument and (regarding microscopies) from the resolution of the images. Estimated file sizes from some KB to few MB (images) per measurement. Sensors evaluation tests: - Revisable. - Files of raw data acquired by dedicated software. ASCII files (in csv format) of approximately some MB per study depending on testing duration and sampling rate with an overall estimated size 50 GB.
	Data Security and Storage	Supplementary and raw data initially handwritten as text in lab books are not digitized but the information is transferred to digital reports stored in office computers. Experimental data concerning physicochemical analyses of membranes are stored in the hard drives of the corresponding instrument(s). Raw data collected during sensors evaluation tests are stored as .csv or spreadsheet (excel) files in lab computers. Reports, documents, photos, designs and other digitized material generated during – and relative to - intelWATT project are stored in office computers as well as in EmDESK platform. According to the lab policy all data are backed up regularly and the backups are stored in an internal network drive as well as removable hard drives. All the computers are protected by hard passwords, firewalls and antivirus / antimalware software. The current protocols prevent access to any external user.
	Data utility	The data are useful for the whole team of POLITO, to all members of the intelWATT project as well as to other research groups focusing in membrane science/technology and relevant applications. End users in the electric power generation and renewable energy fields, water and wastewater treatment professionals, process engineers and post graduate students focusing in the activities of intelWATT are indicative additional groups which benefit from the project's results. The data can be further exploited as source for new funding or new internal scientific projects as well as for commercialization purposes.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 7 – Experimental – WP2,3,4,8 – PARTNER: POLITO		
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Discoverability of the data is achieved by making the analyzed data available (provided that rules and provisions of signed GA, CA and IPR interests are not violated) for publication in scientific journals. Discoverability is further maximized linking the article with the relevant dataset (including related metadata) which is deposited in a scientific data repository. These data can be further exploited as source for new funding or new internal scientific projects. The relative metadata are stored mainly in .txt, and .pdf format. Experimental keywords will be used to create the corresponding metadata. These will include some common fields like the name/code of the material, partner's acronym, as well as some specific info depending on the technique. For example, for membrane evaluation the following fields will be included: Type of experiment, membrane preparation procedure, salinity gradient concentration, etc. Even though the metadata file is not created by a standard schema and is in the form of a simple .txt file, it should provide the following information: Title - Name of the dataset or research project that produced it Creator - Names and addresses of the organization or people who created the data Identifier - Number used to identify the data. Subject - Keywords or phrases describing the subject or content of the data Access information - Where and how your data can be accessed by other researchers Date - Key dates associated with the data, including time period covered by the data Methodology - How the data was generated, including equipment or software used, experimental protocol, other things one might include in a lab notebook Data processing - Information on how the data has been altered or processed.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 7 – Experimental – WP2,3,4,8 – PARTNER: POLITO		
		Sources - Citations to material for data derived from other sources, including details of where the source data is held and how it was accessed. List of file names - List of all data files associated with the dataset, with their names and file extensions. File Formats - Format(s) of the data, e.g. FITS, SPSS, HTML, JPEG, and any software required to read the data. File structure - Organization of the data file(s) and the layout of the variables, when applicable. Variable list - List and description of variables in the data files, when applicable. Code lists - Explanation of codes or abbreviations used in either the file names or the variables in the data files (e.g. '999 indicates a missing value in the data').
	Identifiability of data (refer to standard id mechanisms)	Data is made identifiable by manually naming data files according to the rules of the lab and the file naming conventions referred below. No standard ID mechanism is used. Data (in the form of document, spreadsheets, images, presentation etc.) uploaded to an online data repository is identified by a Digital Object Identification (DOI) which assigned automatically by the repository. The DOI also contains metadata that provides the relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored.
	Naming conventions used	All the results and data generated during the implementation period of the project will be stored in dedicated folders. The folders' structure is tree like. The root folder (named with the project acronym "intelWATT") contains the WP folders. Further structuring inside the WP folders will be based on Task or Case study numbering. For example folder ..\\intelWATT\WP_2\TASK_3 contains files and data relative to the task 3 of WP 2. Data files are manually named in a way that their source can be uniquely distinguished. In addition, the raw data files should be linked with relevant records (in physical form in a lab book or in a text file) containing all the necessary details for the reusability / interoperability of the data (e.g. the experimental set up and conditions, the preparation method of the materials, pre and / or post treatments of the samples etc.). *Second option* *Data are manually named in a way that one can distinguish their source. Also the name of the interpreter/analyser is added.* *These data are used for the preparation of internal reports. The reports are named by the project and the period, in which the data belongs.*

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 7 – Experimental – WP2,3,4,8 – PARTNER: POLITO		
	Search keywords approach	In some cases some keywords are used including the name of the sample and the corresponding technique used. *Second option* *No specific approach is provided. The pattern for keywords generation is mainly based on the kind of experiments performed, the experimental procedures and conditions, characteristics and properties of the under testing materials, analytical methods etc.*
	Clear versioning approach	Revised datasets will take the extension "rev" followed by the number of the revision made. *Second option* *The versioning of data files is manual indicating the date of its creation and the writer.*
	Standards or procedures for metadata creation applied	N/A Please select between N/A and the text below Specific templates will be prepared for the collection of all the necessary information related to a particular dataset. A supplementary .pdf or .txt file will be created using the above information and a standardized layout / format. This file will be uploaded in a data repository where a DOI will be automatically assigned to it. The DOI contains metadata that provides users with relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored. Even though the project discipline is not related directly to any metadata standard or schema, however there are some general purpose standards like the Dublin Core Metadata Initiative that can be applied. Metadata Editors (preferably open access) will be used to generate partly or complete the relevant metadata files.
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	N/A
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	The raw data obtained by the POLITO team belongs to the POLITO. The research team will define the way and the time that these data can be presented. The analyzed data that are presented to the consortium are available for the whole consortium for the full duration of the project and according to the rules and provisions of the signed G.A., C.A. and intellectual property obligations. Accessibility of data out of the scope of intelWATT project requires the permission of POLITO. The same occurs also for the raw data. In such a case the raw data can be made available only upon request and under separate agreement between the partners and as long this request does not interfere with future publication and intellectual property interests, or precede data analysis.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 7 – Experimental – WP2,3,4,8 – PARTNER: POLITO		
	How data will be made available	Data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Disclosable data will be deposited in an open data repository, such as Zenodo or other recommended repositories (see below) provided that rules and provisions in signed GA and CA, IPR obligations or/and confidentiality restrictions are not violated. An option being considered is to link the data and metadata to forthcoming publications. Specific data can be made available upon requests (see 2.4).
	Methods or SW tools for data access	Tools for accessing the data are any text editor and spreadsheet packages for the output files. Optionally photo/image viewers / editors or CAD software may be needed for the visualization of some engineering designs.
	SW documentation and other information needed	In some cases, the supplementary notes may be needed for better understanding of the data.
	Repository for deposit of data, metadata, documentation and code	Specific repositories that will be considered include: FigShare, Mendeley Data, Open Science Framework, Harvard Dataverse, Dryad Digital Repository, Science Data Bank, Zenodo, which can be used for storage of data and metadata. These generalist repositories may also be appropriate for archiving associated analyses, or experimental-control data. *Second Option* *N/A*
	Access restrictions	As a rule, the availability of the data is according to the rule and the provisions of the signed G.A. and C.A., especially to the IPR clauses indicated there. In some circumstances, POLITO will not be able to disclose some of this data until a certain time within the project timeframe due to possible commercial confidentiality and intellectual property reasons.
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A
	Standard vocabulary or mapping to commonly used ontologies	N/A
	Information Portability and exporting	N/A
2.4 FAIR DATA – Increase data re-	**Data licensing for wide reuse**	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 7 – Experimental – WP2,3,4,8 – PARTNER: POLITO		
use (through clarifying licenses)	**Timing of data availability for re-use (incl. indications on embargo)**	Data can be made available after publication in scientific papers. The raw data will be not available to other partners, but can be shared only after permission of the POLITO team.
	Data usability by Third Parties (after the end of the project)	Except for the case where results have been published, data could be reused only with the permission of POLITO and according to the rules established in the G.A. and C.A.
	Restrictions to data re-use	Restrictions to the re-use of data from the Consortium Partners as imposed by the signed G.A. and C.A. rules, specific IPR established agreements and/or possible commercial confidentialities
	Quality assurance process	Data quality is assured through the : - Regular checking and maintenance of the experimental and mechanical equipment including electronic sensors and control devices, - Calibration of sensors, - Calibration and performance verification of analytical instruments, - Use of standards for certain measurements, - Blank experiments and repetition of experiments, - Data entry validation and peer review of data.
	Length of time of data re-usability	N/A
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	N/A
	Data Management Responsibilities	The POLITO team is responsible for the management of the data obtained within the frame of intelWATT project. Researchers of POLITO are high educated and qualified to ensure that dataset is accurate and reliable.
	Long Term Preservation assessment	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

4.1.8 DATA SET n. 8 – Experimental – WP3,4,5,6,7 – PARTNER: CUT

DATA SET n. 8 – Experimental – WP3,4,5,6,7 – PARTNER: CUT		
1 DATA SUMMARY	**Purpose of the Data**	Data collected relate to the development and optimization of membranes and modules. Polymer recipes and manufacturing parameters are recorded, processed, compared and saved. Furthermore, additional data is generated for performance tests in the laboratory, technical and pilot scale. All the generated data are in close relation with the implementation of the project's deliverables and objectives.
	Type and Format of data	▪ Text: field or laboratory notes, communication– in plain text, (txt), HTML, XLM, PDF/A, ▪ Numeric: tables, counts, measurements – in .XLSX, .CSV (performance evaluation tests), ▪ Audiovisual: images, video – in .JPEG, .JPG, .PNG, .TIFF, AIFF, WAVE, .MP3, .MP4 (microscopic techniques SEM, light microscope), ▪ Instrument specific: equipment outputs (specify equipment and format) {physicochemical characterization and analysis - can be converted in XLS and CSV format}.
	Reused-Data	The majority of the data will be new, generated during the implementation of the project.
	Data origin	▪ Observational – data captured in real time (performance evaluation tests, material preparation), usually irreplaceable - i.e. sensor readings, images, telemetries, sample data. The data are reproducible if the same conditions and equipment are used. ▪ Experimental – data from lab equipment (physicochemical analysis, materials characterization), often reproducible - i.e. chromatogram). The observational and the experimental data are subject to analysis. In some case the analysis is very time consuming which increases the cost.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 8 – Experimental – WP3,4,5,6,7 – PARTNER: CUT		
	Data size	Physicochemical analyses: - Revisable, - Data file size depending on the analytical technique and instrument. The Datasets vary from some Kbytes to few Mbytes per experiment. Material characterization: - Revisable, - Collection of data files from various laboratory instruments like SEM, contact angle, viscosimeter, light microscope, pressure hold tests, cut off by GPC. Data files size depends on the instrument and (regarding microscopies) from the resolution of the images. Estimated file sizes from some KB to few MB (images) per measurement. Performance evaluation tests: - Revisable, - Files of raw data acquired by the systems of experimental devises, controllers, online sensors and analyzers etc, cut off measurements by GPC, - Data files depends on the instrument.
	Data Security and Storage	Supplementary and raw data initially handwritten as text in lab books are not digitized but the information is transferred to digital reports stored on the server of the company. Experimental data concerning physicochemical analyses of water samples or characterization of materials (e.g. membranes) are stored in the hard drives of the corresponding instrument(s) and the server of the company. Reports, documents, photos, designs and other digitized material generated during – and relative to - intelWATT project are stored on the server of the company. According to the company policy all data are backed up regularly and the backups are stored in hard drives. All the computers are protected by hard passwords, firewalls and antivirus / antimalware software. The current protocols prevent access to any external user.
	Data utility	The data are useful for the whole team of CUT Membrane Technology, to all members of the intelWATT project. End users in the electric power generation and renewable energy fields, water and wastewater treatment professionals, process engineers and post graduate students focusing in the activities of intelWATT are indicative additional groups which benefit from the project's results. The data can be further exploited as source for commercialization purposes.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 8 – Experimental – WP3,4,5,6,7 – PARTNER: CUT		
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Discoverability of the data is achieved by making the analyzed data available (provided that rules and provisions of signed GA, CA and IPR interests are not violated) for publication in scientific journals. Discoverability is further maximized linking the article with the relevant dataset (including related metadata) which is deposited in a scientific data repository. These data can be further exploited as source for new funding or new internal scientific projects. The relative metadata are stored mainly in .txt, and .pdf format. Experimental keywords will be used to create the corresponding metadata. These will include some common fields like the name/code of the material (membrane or IX resin), partner's acronym, as well as some specific info depending on the technique. For example, for membrane evaluation the following fields will be included: Type of experiment, Feed composition, Temperature, Pressure, Flow rates, Sample info, etc. Even though the metadata file is not created by a standard schema and is in the form of a simple .txt file, it should provide the following information: Title - Name of the dataset or research project that produced it. Creator - Names and addresses of the organization or people who created the data Identifier - Number used to identify the data Subject - Keywords or phrases describing the subject or content of the data Access information - Where and how your data can be accessed by other researchers. Date - Key dates associated with the data, including time period covered by the data Methodology - How the data was generated, including equipment or software used, experimental protocol, other things one might include in a lab notebook Data processing - Information on how the data has been altered or processed Sources - Citations to material for data derived from other sources, including details of where the source data is held and how it was accessed List of file names - List of all data files associated with the dataset, with their names and file extensions File Formats - Format(s) of the data, e.g. FITS, SPSS, HTML, JPEG, and any software required to read the data. File structure - Organization of the data file(s) and the layout of the variables, when applicable. Variable list - List and description of variables in the data files, when applicable Code lists - Explanation of codes or abbreviations used in either the file names or the variables in the data files (e.g. '999 indicates a missing value in the data')

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 8 – Experimental – WP3,4,5,6,7 – PARTNER: CUT		
	Identifiability of data (refer to standard id mechanisms)	Data is made identifiable by manually naming data files according to the rules of the lab and the file naming conventions referred below. No standard ID mechanism is used. Data (in the form of document, spreadsheets, images, presentation etc.) uploaded to an online data repository is identified by a Digital Object Identification (DOI) which assigned automatically by the repository. The DOI also contains metadata that provides the relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored.
	Naming conventions used	All the results and data generated during the implementation period of the project will be stored in dedicated folders. The folders' structure is tree like. The root folder (named with the project acronym "intelWATT") contains several sub folders. Further structuring inside the sub folders will be based on Task or Case study numbering. Data files are manually named in a way that their source can be uniquely distinguished. In addition, the raw data files should be linked with relevant records (in physical form in a lab book or in a text file) containing all the necessary details for the reusability / interoperability of the data (e.g. the experimental set up and conditions, the preparation method of the materials, pre and / or post treatments of the samples etc.).
	Search keywords approach	N/A
	Clear versioning approach	The versioning of data files is manual indicating the date of its creation and the writer.
	Standards or procedures for metadata creation applied	N/A
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	N/A
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	The raw data obtained by the CUT team belongs to the CUT. The research team will define the way and the time that these data can be presented. The analyzed data that are presented to the consortium are available for the whole consortium for the full duration of the project and according to the rules and provisions of the signed G.A., C.A. and intellectual property obligations. Accessibility of data out of the scope of intelWATT project requires the permission of CUT. The same occurs also for the raw data.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 8 – Experimental – WP3,4,5,6,7 – PARTNER: CUT		
	How data will be made available	Data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Specific data can be made available upon requests (see 2.4).
	Methods or SW tools for data access	Tools for accessing the data are any text editor and spreadsheet packages for the output files. Optionally photo/image viewers / editors or CAD software may be needed for the visualization of some engineering designs.
	SW documentation and other information needed	In some cases, the supplementary notes may be needed for better understanding of the data.
	Repository for deposit of data, metadata, documentation and code	N/A
	Access restrictions	As a rule, the availability of the data is according to the rule and the provisions of the signed G.A. and C.A., especially to the IPR clauses indicated there. In some circumstances, CUT will not be able to disclose some of this data until a certain time within the project timeframe due to possible commercial confidentiality and intellectual property reasons.
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A
	Standard vocabulary or mapping to commonly used ontologies	N/A
	Information Portability and exporting	N/A
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)	**Data licensing for wide reuse**	N/A
	Timing of data availability for re-use (incl. indications on embargo)	Data can be made available after publication in scientific papers. The raw data will be not available to other partners, but can be shared only after permission of CUT.
	Data usability by Third Parties (after the end of the project)	Except for the case where results have been published, data could be reused only with the permission of CUT and according to the rules established in the G.A. and C.A.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 8 – Experimental – WP3,4,5,6,7 – PARTNER: CUT		
	Restrictions to data re-use	Restrictions to the re-use of data from the Consortium Partners as imposed by the signed G.A. and C.A. rules, specific IPR established agreements and/or possible commercial confidentialities.
	Quality assurance process	Data quality is assured through the : - Regular checking and maintenance of the experimental and mechanical equipment including electronic sensors and control devices, - Calibration of sensors (temperature, pressure, flow, conductivity etc.), - Calibration and performance verification of analytical instruments, - Use of standards for certain measurements, - Blank experiments and repetition of experiments, - Data entry validation and peer review of data.
	Length of time of data re-usability	N/A
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	N/A
	Data Management Responsibilities	The CUT team is responsible for the management of the data obtained within the frame of intelWATT project. Researchers of CUT are high educated and qualified to ensure that dataset is accurate and reliable.
	Long Term Preservation assessment	Not N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

4.1.9 DATA SET n. 9 – Experimental – WP2,3,4,7 – PARTNER: BIA

DATA SET n. 9 – Experimental– WP2,3,4,7 – PARTNER: BIA		
1 DATA SUMMARY	**Purpose of the Data**	Data concerns physicochemical analyses of relevant water streams for the case study No 3 (as indicated and supplied by Partner No 10, BIA Kunststoff- und Galvanotechnik GmbH & Co. KG, material characterization (such as ion exchange resins or activated carbon), membrane module design development and performance evaluation studies (relative to process conditions, technologies combinations, modules configurations and setups, membrane shape and dimension etc.) at lab and pilot scale, process design drawing(s). All the generated data are in close relation with the implementation of the project's deliverables and objectives.
	Type and Format of data	▪ Text: field or laboratory notes, survey responses – in plain text, (txt), HTML, XLM, PDF/A {performance evaluation tests and material preparation}, ▪ Numeric: tables, counts, measurements – in .XLSX, .CSV (performance evaluation tests), ▪ Audiovisual: images, video – in .JPEG, .JPG, .PNG, .TIFF, AIFF, WAVE, .MP3, .MP4 (microscopic techniques SEM, AFM). ▪ Instrument specific: equipment outputs (specify equipment and format) {physicochemical characterization and analysis - can be converted in XLS, ASCII and CSV format}.
	Reused-Data	The majority of the data will be new, generated during the implementation of the project.
	Data origin	▪ Observational – data captured in real time (performance evaluation tests, material preparation), usually irreplaceable - i.e. sensor readings, images, telemetries, sample data. The data are reproducible if the same conditions and equipment are used. ▪ Experimental – data from lab equipment (physicochemical analysis, materials characterization), often reproducible - i.e. chromatograms, absorbance spectra, x-ray diffraction data, porometer, porosimeter. The observational and the experimental data are subject to analysis. In some cases, the analysis is very time consuming which increases the cost.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 9 – Experimental– WP2,3,4,7 – PARTNER: BIA		
	Data size	Physicochemical analyses: - Revisable. - Data file size depending on the analytical technique and instrument. The Datasets vary from some Kbytes (AAS) to few Mbytes (Ion chromatography) per experiment. Material characterization: - Revisable, - Collection of data files from various laboratory instruments like SEM, AFM, contact angle, surface tension etc. Data files size depends on the instrument and (regarding microscopies) from the resolution of the images. Estimated file sizes from some KB to few MB (images) per measurement. Performance evaluation tests: - Revisable, - Files of raw data acquired by the SCADA systems of experimental devises, controllers, online sensors and analyzers etc. ASCII files (in csv format) of approximately 10 MB per study depending on testing duration and sampling rate with an overall estimated size 50 GB , - The performance evaluation provides data that can be amended. In this case the test should be repeated in order to obtain new data.
	Data Security and Storage	Supplementary and raw data initially handwritten as text in lab books are not digitized but the information is transferred to digital reports stored in office computers. Experimental data concerning physicochemical analyses of water samples or characterization of materials (e.g. membranes) are stored in the hard drives of the corresponding instrument(s). Raw data collected during performance evaluation tests by SCADA systems of experimental devices are stored as .csv or spreadsheet (excel) files in lab computers. Reports, documents, photos, designs and other digitized material generated during – and relative to - intelWATT project are stored in office computers as well as in EmDESK platform. According to the lab policy all data are backed up regularly and the backups are stored in an internal network drive as well as removable hard drives. All the computers are protected by hard passwords, firewalls and antivirus / antimalware software. The current protocols prevent access to any external user.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 9 – Experimental– WP2,3,4,7 – PARTNER: BIA		
	Data utility	The data are useful for the whole team of BIA, to all members of the intelWATT project as well as to other research groups focusing in membrane science/technology and relevant applications. End users in the electric power generation and renewable energy fields, water and wastewater treatment professionals, process engineers and post graduate students focusing in the activities of intelWATT are indicative additional groups which benefit from the project's results. The data can be further exploited as source for new funding or new internal scientific projects as well as for commercialization purposes.
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Discoverability of the data is achieved by making the analyzed data available (provided that rules and provisions of signed GA, CA and IPR interests are not violated) for publication in scientific journals. Discoverability is further maximized linking the article with the relevant dataset (including related metadata) which is deposited in a scientific data repository. These data can be further exploited as source for new funding or new internal scientific projects. The relative metadata are stored mainly in .txt, and .pdf format. Experimental keywords will be used to create the corresponding metadata. These will include some common fields like the name/code of the material (membrane or IX resin), partner's acronym, as well as some specific info depending on the technique. For example, for membrane evaluation the following fields will be included: Type of experiment, Feed composition, Temperature, Pressure, Flow rates, Sample info, etc. Even though the metadata file is not created by a standard schema and is in the form of a simple .txt file, it should provide the following information: Title - Name of the dataset or research project that produced it. Creator - Names and addresses of the organization or people who created the data Identifier - Number used to identify the data. Subject - Keywords or phrases describing the subject or content of the data Access information - Where and how your data can be accessed by other researchers Date - Key dates associated with the data, including time period covered by the data Methodology - How the data was generated, including equipment or software used, experimental protocol, other things one might include in a lab notebook Data processing - Information on how the data has been altered or processed

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 9 – Experimental– WP2,3,4,7 – PARTNER: BIA		
		Sources - Citations to material for data derived from other sources, including details of where the source data is held and how it was accessed. List of file names - List of all data files associated with the dataset, with their names and file extensions. File Formats - Format(s) of the data, e.g. FITS, SPSS, HTML, JPEG, and any software required to read the data. File structure - Organization of the data file(s) and the layout of the variables, when applicable. Variable list - List and description of variables in the data files, when applicable. Code lists - Explanation of codes or abbreviations used in either the file names or the variables in the data files (e.g. '999 indicates a missing value in the data').
	Identifiability of data (refer to standard id mechanisms)	Data is made identifiable by manually naming data files according to the rules of the lab and the file naming conventions referred below. No standard ID mechanism is used. Data (in the form of document, spreadsheets, images, presentation etc.) uploaded to an online data repository is identified by a Digital Object Identification (DOI) which assigned automatically by the repository. The DOI also contains metadata that provides the relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored.
	Naming conventions used	All the results and data generated during the implementation period of the project will be stored in dedicated folders. The folders' structure is tree like. The root folder (named with the project acronym "intelWATT") contains the WP folders. Further structuring inside the WP folders will be based on Task or Case study numbering. For example folder ..\\intelWATT\WP_2\TASK_3 contains files and data relative to the task 3 of WP 2. Data files are manually named in a way that their source can be uniquely distinguished. For example, a name consisting of a prefix (e.g. characteristic of the experimental technique or material) followed by a date and/or time stamp could be an appropriate convention for a file containing raw data from an experimental device's SCADA systems. n addition, the raw data files should be linked with relevant records (in physical form in a lab book or in a text file) containing all the necessary details for the reusability / interoperability of the data (e.g. the experimental set up and conditions, the preparation method of the materials, pre and / or post treatments of the samples etc.).
	Search keywords approach	In some cases some keywords are used including the name of the sample and the corresponding technique used.
	Clear versioning approach	Revised datasets will take the extension "rev" followed by the number of the revision made.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 9 – Experimental– WP2,3,4,7 – PARTNER: BIA		
	Standards or procedures for metadata creation applied	N/A
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	N/A
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	The raw data obtained by the BIA team belongs to BIA. The research team will define the way and the time that these data can be presented. The analyzed data that are presented to the consortium are available for the whole consortium for the full duration of the project and according to the rules and provisions of the signed G.A., C.A. and intellectual property obligations. Accessibility of data out of the scope of intelWATT project requires the permission of BIA. The same occurs also for the raw data. In such a case the raw data can be made available only upon request and under separate agreement between the partners and as long this request does not interfere with future publication and intellectual property interests, or precede data analysis.
	How data will be made available	Data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Disclosable data will be deposited in an open data repository, such as Zenodo or other recommended repositories (see below) provided that rules and provisions in signed GA and CA, IPR obligations or/and confidentiality restrictions are not violated. An option being considered is to link the data and metadata to forthcoming publications. Specific data can be made available upon requests (see 2.4)
	Methods or SW tools for data access	Tools for accessing the data are any text editor and spreadsheet packages for the output files. Optionally photo/image viewers / editors or CAD software may be necessary for the visualization of some engineering designs.
	SW documentation and other information needed	In some cases, the supplementary notes may be necessary for better understanding of the data.
	Repository for deposit of data, metadata, documentation and code	Specific repositories that will be considered include: FigShare, Mendeley Data, Open Science Framework, Harvard Dataverse, Dryad Digital Repository, Science Data Bank, Zenodo, which can be used for storage of data and metadata. These generalist repositories may also be appropriate for archiving associated analyses, or experimental-control data.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 9 – Experimental– WP2,3,4,7 – PARTNER: BIA		
	Access restrictions	As a rule, the availability of the data is according to the rule and the provisions of the signed G.A. and C.A., especially to the IPR clauses indicated there. In some circumstances, BIA will not be able to disclose some of this data until a certain time within the project timeframe due to possible commercial confidentiality and intellectual property reasons.
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A
	Standard vocabulary or mapping to commonly used ontologies	N/A
	Information Portability and exporting	N/A
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)	**Data licensing for wide reuse**	N/A
	Timing of data availability for re-use (incl. indications on embargo)	Data can be made available after publication in scientific papers. The raw data will be not available to other partners, but can be shared only after permission of the BIA team.
	Data usability by Third Parties (after the end of the project)	Except for the case where results have been published, data could be reused only with the permission of BIA and according to the rules established in the G.A. and C.A.
	Restrictions to data re-use	Restrictions to the re-use of data from the Consortium Partners as imposed by the signed G.A. and C.A. rules, specific IPR established agreements and/or possible commercial confidentialities.
	Quality assurance process	Data quality is assured through the : - Regular checking and maintenance of the experimental and mechanical equipment including electronic sensors and control devices, - Calibration of sensors (temperature, pressure, flow, conductivity etc.), - Calibration and performance verification of analytical instruments, - Use of standards for certain measurements, - Blank experiments and repetition of experiments, - Data entry validation and peer review of data.
	Length of time of data re-usability	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 9 – Experimental– WP2,3,4,7 – PARTNER: BIA		
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	N/A
	Data Management Responsibilities	The BIA team is responsible for the management of the data obtained within the frame of intelWATT project. Researchers of BIA are high educated and qualified to ensure that dataset is accurate and reliable.
	Long Term Preservation assessment	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

4.1.10 DATA SET n. 10 – Experimental – WP2, 5, – PARTNER: IHE DELFT

DATA SET n. 10 – Experimental – WP2, 5, – PARTNER: IHE DELFT		

1 DATA SUMMARY	**Purpose of the Data**	Data concerns analysis of membrane fouling/scaling related water quality parameters of relevant water streams for the case studies, membrane autopsies of fouled membranes used in pilot as well as lab scale plants.
		Following water quality parameters will be analysed;
		- Silt density index (SDI)
		- Modified fouling index ($MFI_{0.45}$)
		- Modified Fouling Index (MFI – UF)
		- Transparent Exopolymer Particles (TEPs)
		- Biological/Organic fouling
		  - Bacterial growth potential (BGP)
		  - Liquid Chromatography – Organic Carbon Detection (LC-OCD), Total Nitrogen (TN), and Orthophosphate
		- Scaling prediction in membrane systems using software
		  - Genesys Membrane Master (MM4) – Genesys International
		  - WAVE-DuPont membrane projection software
		  - IMSDesign– Hydranautics membrane projection software
		All the generated data are in close relation with the implementation of the project's deliverables and objectives.

	Type and Format of data	▪ Text: field or laboratory notes, (txt), PDF/A, ▪ Numeric: tables, counts, measurements – in .XLSX, ▪ Instrument specific: equipment outputs (specify equipment and format) {water quality analysis - can be converted in XLS, and CSV format},
	Reused-Data	The data will be new, generated during the implementation of the project.
	Data origin	▪ Experimental – data from lab equipment (water quality data analysis, materials characterization). ▪ Simulation – data from running models e.g., membrane scaling simulation model
	Data size	Water quality analysis - Revisable. - Data file size depending on the analytical technique and instrument. The Datasets vary from some Kbytes (AAS) to few Mbytes per experiment. Scaling modelling - Revisable. - Collection of data files from modelling software. Data files size from some KB to few MB per modelling.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 10 – Experimental – WP2, 5, – PARTNER: IHE DELFT		
	Data Security and Storage	All the data generated will be stored in office computer, in EmDESK platform as well as backed up in external hard drive All the computers are protected by hard passwords, firewalls and antivirus / antimalware software. The current protocols prevent access to any external user.
	Data utility	The data are useful for the whole team of IHE Delft, to all members of the intelWATT project as well as to other research groups focusing in membrane technology and relevant applications.
2.1 FAIR DATA - Making data findable	Discoverability of data (metadata provision)	Discoverability of the data is achieved by making the analyzed data available (provided that rules and provisions of signed GA, CA and IPR interests are not violated) for publication in scientific journals. Sources - Citations to material for data derived from other sources, including details of where the source data is held and how it was accessed List of file names - List of all data files associated with the dataset, with their names and file extensions Variable list - List and description of variables in the data files, when applicable
	Identifiability of data (refer to standard id mechanisms)	N/A
	Naming conventions used	All the results and data generated during the implementation period of the project will be stored in dedicated folders. For example folder ..\\intelWATT\WP_2\TASK_2.6 contains files and data relative to the task 2.6 of WP 2.
	Search keywords approach	No specific approach is provided.
	Clear versioning approach	N/A
	Standards or procedures for metadata creation applied	N/A
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	N/A
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	The analyzed data/report that are presented to the consortium are available for the whole consortium for the full duration of the project and according to the rules and provisions of the signed G.A., C.A. and intellectual property obligations.
	How data will be made available	Data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Data can be available in the form of report to consortium partners provided that rules and provisions in signed GA and CA,

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 10 – Experimental – WP2, 5, – PARTNER: IHE DELFT		
	Methods or SW tools for data access	N/A
	SW documentation and other information needed	N/A
	Repository for deposit of data, metadata, documentation and code	N/A
	Access restrictions	N/A
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A
	Standard vocabulary or mapping to commonly used ontologies	N/A
	Information Portability and exporting	N/A
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)	**Data licensing for wide reuse**	N/A
	Timing of data availability for re-use (incl. indications on embargo)	Data can be made available in the form of report.
	Data usability by Third Parties (after the end of the project)	Except for the case where results have been published, data could be reused only with the permission of IHE Delft and according to the rules established in the G.A. and C.A.
	Restrictions to data re-use	According to the signed G.A. and C.A. rules.
	Quality assurance process	Data quality is assured through the : - Use of standard controls for each set of experiments, - Blank experiments, - Duplicate/triplicate of the each dataset of experiment, - Data entry validation and peer review of data.
	Length of time of data re-usability	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 10 – Experimental – WP2, 5, – PARTNER: IHE DELFT		
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	N/A
	Data Management Responsibilities	The IHE team working for intelWATT project is responsible for the management of the data obtained within the frame of intelWATT project.
	Long Term Preservation assessment	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

4.1.11 DATA SET n. 11 – Experimental LCSA Data – WP9 – PARTNER: Studio Fieschi

DATA SET n. 11 – LCSA Data – WP9 – PARTNER: Studio Fieschi		
1 DATA SUMMARY	**Purpose of the Data**	To perform the Life Cycle Sustainability Assessment of the three case studies in WP9 Studio Fieschi will use data collected by other partners (primary data) and from international databases or other publications (secondary data), in order to evaluate the environmental, economic and social impact of the new technologies with a life cycle approach. Data produced by LCSA will be used by the consortium to evaluate the sustainability and the circularity level of the new technologies
	Type and Format of data	Used data: ▪ Numeric: tables, counts, measurements – in .XLSX, .CSV, etc. ▪ Text: processes description, survey responses – in .DOCX, PDF/A, etc. Generated data: ▪ Numeric: tables, graphs – in .XLSX, .CSV , etc.Text: study report – in PDF/A , etc.
	Reused-Data	In the sustainability assessment some generic data can be used, collected from literature and international databases (i.e. Ecoinvent).
	Data origin	Data for the sustainability assessment will be collected from partners and consist with the quantification of energy and material flows for each life cycle stage, economic costs of materials and processes, information on social aspects. Data sources could be observation, experiment, simulation.
	Data size	Dataset is growing and/or revisable; maximum size is about 150 MB
	Data Security and Storage	Data will be in electronic form, stored in computer and in back-up server in the company site and in cloud.
	Data utility	The information on the assessment can be used by partners to evaluate the sustainability performance of new technologies. The assessment methodology could be used to evaluate other kind of technologies
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Data generated in .xlsx format by LCA software include information about data (metadata) and modelling process.
	Identifiability of data (refer to standard id mechanisms)	N/A
	Naming conventions used	File name will refer to project name (IntelWATT), Task, and content.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 11 – LCSA Data – WP9 – PARTNER: Studio Fieschi		
	Search keywords approach	There will be no keywords referred to LCSA results.
	Clear versioning approach	File name contains the number of revision (REV_00).
	Standards or procedures for metadata creation applied	Metadata are generated by LCA software automatically.
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	Data wouldn't be grouped in records organized in hierarchical structure.
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	Data are available to the Partners of IntelWATT project for the whole duration of the project and according to the rules of the signed G.A. and C.A.
	How data will be made available	Data will be available to other partners through the Collaborative platform used in the project (EMDESK).
	Methods or SW tools for data access	N/A
	SW documentation and other information needed	N/A
	Repository for deposit of data, metadata, documentation and code	N/A
	Access restrictions	No assess restriction provided
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A
	Standard vocabulary or mapping to commonly used ontologies	N/A
	Information Portability and exporting	N/A
2.4 FAIR DATA – Increase data re-	**Data licensing for wide reuse**	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 11 – LCSA Data – WP9 – PARTNER: Studio Fieschi		
use (through clarifying licenses)	**Timing of data availability for re-use (incl. indications on embargo)**	N/A
	Data usability by Third Parties (after the end of the project)	Data can be reused only with the permission of Studio Fieschi.
	Restrictions to data re-use	Restriction on data re-use is imposed by Intellectual Property; in re-using data their origin must be declared.
	Quality assurance process	Data quality is assured by data entry validation and peer review of data.
	Length of time of data re-usability	No time limit provided.
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	Additional costs are not expected.
	Data Management Responsibilities	Studio Fieschi & soci is responsible for the management of the data obtained within the project.
	Long Term Preservation assessment	N/A

4.1.12 DATA SET n. 12 – Theoretical – WP2,3,4,5,6,8– PARTNER: TECHEDGE

DATA SET n. 12 – Theoretical – WP2,3,4,5,6,8 – PARTNER: TECHEDGE		
1 DATA SUMMARY	**Purpose of the Data**	Sensor data, computer SW, general project documents.
	Type and Format of data	Describe the type of data used or generated within the project, specifying the form and format of the data: ▪ Computer code (Python- notebook), ▪ Sensor data CSV, TXT, ▪ General Office docs project architecture,
	Reused-Data	There is not reused data.
	Data origin	▪ Experimental – data from membrane labs scale(sensors), ▪ Simulation – data generated by ML Model, ▪ Derived/Compiled – data output from membrane simulator and from ML model,

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 12 – Theoretical – WP2,3,4,5,6,8 – PARTNER: TECHEDGE		
	Data size	▪ Growing: new data may be added, but the old data is never changed or deleted. We estimate 2GbB of data pers use Case
	Data Security and Storage	Data is stored in gDrive o project Datalake (AWS)
	Data utility	WP8 member will use the data as well like will use it together with WP2 and use cases members.
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Search by metadata (date, model, sensor).
	Identifiability of data (refer to standard id mechanisms)	N/A
	Naming conventions used	[USE-CASE, SENSOR NAME, DATE, TIME]
	Search keywords approach	Tagging.
	Clear versioning approach	Versioning only on software code.
	Standards or procedures for metadata creation applied	Code version for sw [USE-CASE, SENSOR NAME, DATE, TIME].
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	No
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	Data will be available to all consortium member other data to be shared must be discussed.
	How data will be made available	Repository for a specific user group (password protected).
	Methods or SW tools for data access	None
	SW documentation and other information needed	Notebook format will provide sw code documentation. Other architectural documentation will be provided in ppt / word format.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 12 – Theoretical – WP2,3,4,5,6,8 – PARTNER: TECHEDGE		
	Repository for deposit of data, metadata, documentation and code	Project DataLake Code: It is not planned for the moment.
	Access restrictions	It is not planned for the moment.
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	All Data will be accessible via API with authentication.
	Standard vocabulary or mapping to commonly used ontologies	No
	Information Portability and exporting	CSV, TXT sensor data should be exported in local pc.
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)	**Data licensing for wide reuse**	It is not planned for the moment.
	Timing of data availability for re-use (incl. indications on embargo)	It is not planned for the moment.
	Data usability by Third Parties (after the end of the project)	It is not planned for the moment.
	Restrictions to data re-use	It is not planned for the moment.
	Quality assurance process	Must be agreed and discussed by WP8 members
	Length of time of data re-usability	It is not planned for the moment
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	…
	Data Management Responsibilities	Techedge will be managed the data lake in which data sensors are stored.
	Long Term Preservation assessment	It is not planned at the moment.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

4.1.13DATA SET n. 13 – Experimental – WP2, 3, 6 – PARTNER: ACSA

DATA SET n. 13 – Experimental – WP2, 3, 6 – PARTNER: ACSA		
1 DATA SUMMARY	**Purpose of the Data**	Data concerns physicochemical analyses of relevant water streams for the case study No2 (CS2), characterization of low and high salinity streams (WP2), analysis and characterization of the main experimental parameters regarding inlet and outlet streams from CS2 at the lab scale (WP3) and pilot scale (WP6). All the generated data are in close relation with the implementation of the project's deliverables and objectives.
	Type and Format of data	▪ Text: field or laboratory notes, survey responses – in plain text, (txt), HTML, XLM, PDF/A {performance evaluation tests and material preparation} , ▪ Numeric: tables, counts, measurements – in .XLSX, .CSV (performance evaluation tests), ▪ Audiovisual: images, video – in .JPEG, .JPG, .PNG, .TIFF, AIFF, WAVE, .MP3, ▪ Instrument specific: equipment outputs (specify equipment and format) {physicochemical characterization and analysis - can be converted in XLS, ASCII and CSV format}.
	Reused-Data	The majority of the data will be new, generated during the implementation of the project.
	Data origin	▪ Observational – data captured in real time (performance evaluation tests, material preparation), usually irreplaceable - i.e. sensor readings, images, telemetries, sample data. The data are reproducible if the same conditions and equipment are used, ▪ Experimental – data from lab equipment (physicochemical analysis, water characterization), often reproducible - i.e. chromatograms, absorbance spectra, x-ray diffraction data, ICP-MS, spectrophotometry, gravimetry. The observational and the experimental data are subject to analysis. In some case the analysis is very time consuming which increases the cost.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 13 – Experimental – WP2, 3, 6 – PARTNER: ACSA		
	Data size	Physicochemical analyses: - Revisable, - Data file size depending on the analytical technique and instrument. The Datasets vary from some Kbytes (AAS) to few Mbytes (Ion chromatography) per experiment. Material characterization: - Revisable, - Collection of data files from various laboratory instruments like spectrophotometry, gravimetry. Data files size depends on the instrument. Estimated file sizes from some KB to few MB (images) per measurement. Performance evaluation tests: - Revisable, - Files of raw data acquired by the SCADA systems of experimental devises, controllers, online sensors and analyzers etc. ASCII files (in csv format) of approximately 10 MB per study depending on testing duration and sampling rate with an overall estimated size 50 GB, - The performance evaluation provides data that can be amended. In this case the test should be repeated in order to obtain new data.
	Data Security and Storage	Supplementary and raw data initially handwritten as text in lab books are not digitized but the information is transferred to digital reports stored in office computers. Experimental data concerning physicochemical analyses of water samples or characterization are stored in analytical reports. Raw data collected during performance evaluation tests by SCADA systems of experimental devices are stored as .csv or spreadsheet (excel) files in lab computers. Reports, documents, photos, designs and other digitized material generated during – and relative to - intelWATT project are stored in office computers as well as in EmDESK platform. According to the lab policy all data are backed up regularly and the backups are stored in an internal network drive as well as removable hard drives. All the computers are protected by hard passwords, firewalls and antivirus / antimalware software. The current protocols prevent access to any external user.
	Data utility	The data are useful for the whole team of CS2. End users in the electric power generation and renewable energy fields, water and wastewater treatment professionals, process engineers and post graduate students focusing in the activities of intelWATT are indicative additional groups which benefit from the project's results. The data can be further exploited as source for new funding or new internal scientific projects as well as for commercialization purposes.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 13 – Experimental – WP2, 3, 6 – PARTNER: ACSA		
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Discoverability of the data is achieved by making the analyzed data available (provided that rules and provisions of signed GA, CA and IPR interests are not violated) for publication in scientific journals. Discoverability is further maximized linking the article with the relevant dataset (including related metadata) which is deposited in a scientific data repository. These data can be further exploited as source for new funding or new internal scientific projects. The relative metadata are stored mainly in .txt, and .pdf format. Experimental keywords will be used to create the corresponding metadata. These will include some common fields like the name/code of the sample, partner's acronym, as well as some specific info depending on the technique. For example, for membrane evaluation the following fields will be included: Type of experiment, Feed composition, Temperature, Pressure, Flow rates, Sample info, etc. Even though the metadata file is not created by a standard schema and is in the form of a simple .txt file, it should provide the following information: Title - Name of the dataset or research project that produced it. Creator - Names and addresses of the organization or people who created the data. Identifier - Number used to identify the data. Subject - Keywords or phrases describing the subject or content of the data. Access information - Where and how your data can be accessed by other researchers. Date - Key dates associated with the data, including time period covered by the data Methodology - How the data was generated, including equipment or software used, experimental protocol, other things one might include in a lab notebook. Data processing - Information on how the data has been altered or processed.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 13 – Experimental – WP2, 3, 6 – PARTNER: ACSA		
		Sources - Citations to material for data derived from other sources, including details of where the source data is held and how it was accessed. List of file names - List of all data files associated with the dataset, with their names and file extensions. File Formats - Format(s) of the data, e.g. FITS, SPSS, HTML, JPEG, and any software required to read the data. File structure - Organization of the data file(s) and the layout of the variables, when applicable. Variable list - List and description of variables in the data files, when applicable. Code lists - Explanation of codes or abbreviations used in either the file names or the variables in the data files (e.g. '999 indicates a missing value in the data').
	Identifiability of data (refer to standard id mechanisms)	Data is made identifiable by manually naming data files according to the rules of the lab and the file naming conventions referred below. No standard ID mechanism is used. Data (in the form of document, spreadsheets, images, presentation etc.) uploaded to an online data repository is identified by a Digital Object Identification (DOI) which assigned automatically by the repository. The DOI also contains metadata that provides the relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored.
	Naming conventions used	All the results and data generated during the implementation period of the project will be stored in dedicated folders. The folders' structure is tree like. The root folder (named with the project acronym "intelWATT") contains the WP folders. Further structuring inside the WP folders will be based on Task or Case study numbering. For example folder ..\\intelWATT\WP_2\TASK_3 contains files and data relative to the task 3 of WP 2. Data files are manually named in a way that their source can be uniquely distinguished. For example, a name consisting of a prefix (e.g. characteristic of the experimental technique or material) followed by a date and/or time stamp could be an appropriate convention for a file containing raw data from an experimental device's SCADA systems. In addition, the raw data files should be linked with relevant records (in physical form in a lab book or in a text file) containing all the necessary details for the reusability / interoperability of the data (e.g. the experimental set up and conditions, the preparation method of the materials, pre and / or post treatments of the samples etc.). Data are manually named in a way that one can distinguish their source. Also the name of the interpreter/analyser is added. These data are used for the preparation of internal reports. The reports are named by the project and the period, in which the data belongs.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 13 – Experimental – WP2, 3, 6 – PARTNER: ACSA		
	Search keywords approach	In some cases some keywords are used including the name of the sample and the corresponding technique used. No specific approach is provided. The pattern for keywords generation is mainly based on the kind of experiments performed, the experimental procedures and conditions, characteristics and properties of the under testing materials, analytical methods etc.
	Clear versioning approach	Revised datasets will take the extension "rev" followed by the number of the revision made. The versioning of data files is manual indicating the date of its creation and the writer.
	Standards or procedures for metadata creation applied	N/A
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	N/A
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	The raw data obtained by ACSA team belongs to the ACSA and technology suppliers involved in CS2. The research team will define the way and the time that these data can be presented. The analyzed data that are presented to the consortium are available for the whole consortium for the full duration of the project and according to the rules and provisions of the signed G.A., C.A. and intellectual property obligations. Accessibility of data out of the scope of intelWATT project requires the permission of ACSA. The same occurs also for the raw data. In such a case the raw data can be made available only upon request and under separate agreement between the partners and as long this request does not interfere with future publication and intellectual property interests or precede data analysis.
	How data will be made available	Data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Disclosable data will be deposited in an open data repository, such as Zenodo or other recommended repositories (see below) provided that rules and provisions in signed GA and CA, IPR obligations or/and confidentiality restrictions are not violated. An option being considered is to link the data and metadata to forthcoming publications. Specific data can be made available upon requests (see 2.4).
	Methods or SW tools for data access	Tools for accessing the data are any text editor and spreadsheet packages for the output files. Optionally photo/image viewers / editors or CAD software may be needed for the visualization of some engineering designs.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 13 – Experimental – WP2, 3, 6 – PARTNER: ACSA		
	SW documentation and other information needed	In some cases, the supplementary notes may be needed for better understanding of the data.
	Repository for deposit of data, metadata, documentation and code	Specific repositories that will be considered include: FigShare, Mendeley Data, Open Science Framework, Harvard Dataverse, Dryad Digital Repository, Science Data Bank, Zenodo, which can be used for storage of data and metadata. These generalist repositories may also be appropriate for archiving associated analyses, or experimental-control data.
	Access restrictions	As a rule, the availability of the data is according to the rule and the provisions of the signed G.A. and C.A., especially to the IPR clauses indicated there. In some circumstances, ACSA will not be able to disclose some of this data until a certain time within the project timeframe due to possible commercial confidentiality and intellectual property reasons.
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A
	Standard vocabulary or mapping to commonly used ontologies	N/A
	Information Portability and exporting	N/A
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)	**Data licensing for wide reuse**	N/A
	Timing of data availability for re-use (incl. indications on embargo)	Data can be made available after publication in scientific papers. The raw data will be available to other partners from CS2, but can be shared only after permission of the CS2 team.
	Data usability by Third Parties (after the end of the project)	Except for the case where results have been published, data could be reused according to the rules established in the G.A. and C.A.
	Restrictions to data re-use	Restrictions to the re-use of data from the Consortium Partners as imposed by the signed G.A. and C.A. rules, specific IPR established agreements and/or possible commercial confidentialities

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 13 – Experimental – WP2, 3, 6 – PARTNER: ACSA		
	Quality assurance process	Data quality is assured through the : - Regular checking and maintenance of the experimental and mechanical equipment including electronic sensors and control devices, - Calibration of sensors (temperature, pressure, flow, conductivity etc.), - Calibration and performance verification of analytical instruments, - Use of standards for certain measurements, - Blank experiments and repetition of experiments, - Data entry validation and peer review of data.
	Length of time of data re-usability	N/A
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	N/A
	Data Management Responsibilities	The ACSA team is responsible for the management of the data obtained within the frame of intelWATT project. Researchers of ACSA are high educated and qualified to ensure that dataset is accurate and reliable.
	Long Term Preservation assessment	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

4.1.14DATA SET n. 14 – Replication study – WP9 – PARTNER: UJ

DATA SET n. 14 – Replication study – WP9 – PARTNER: UJ		
1 DATA SUMMARY	**Purpose of the Data**	Data for characteristics of wastewater from different industrial plant for replication potentials study of IntelWATT technologies in Mediterranean region
	Type and Format of data	▪ Text: field or laboratory notes, survey responses – in plain text, (txt), HTML, XLM, PDF/A etc, ▪ Numeric: tables, counts, measurements – in .XLSX, .CSV etc ▪ Audiovisual: images.JPEG, .JPG, .PNG, .TIFF etc.
	Reused-Data	There will be new data from some industrial wastewater to be characterized in our labs, while other data may be reused from published articles or industrial plants own reports.
	Data origin	Experimental- data from lab equipment (physicochemical analysis) Reference or Canonical – collection or conglomeration of smaller (peer-reviewed) datasets probably published and curated - i.e. Published chemical analysis.
	Data size	Physicochemical analysis of industrial wastewater: fixed.
	Data Security and Storage	Wastewater characteristics will be stored in office computer, and all data will be backed up in cloud drive.
	Data utility	The data is useful to investigate replication potentials of the project new technologies in the Mediterranean area, which are known to have water shortage problems.
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Data will be stored as txt or pdf files. Data can be used in publications in scientific journals.
	Identifiability of data (refer to standard id mechanisms)	Data will be made identifiable using naming system that provide information on its source and date.
	Naming conventions used	Naming system will include the wastewater source, the date, the name of person involved in providing the data. All data will be stored in folder of specific work package within intelWATT project folder.
	Search keywords approach	Keywords will depend on water source.
	Clear versioning approach	The versioning of data file is manual indicating the date of creation and the writer.
	Standards or procedures for metadata creation applied	N/A
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 14 – Replication study – WP9 – PARTNER: UJ		
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	Data of industrial wastewater will be accessed by involved partners within the project.
	How data will be made available	Data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Disclosable data will be deposited in an open data repository, provided that rules and provisions in signed GA and CA, IPR obligations or/and confidentiality restrictions are not violated.
	Methods or SW tools for data access	Tools for accessing the data are any text editor and spreadsheet packages for the output files. Optionally photo/image viewers / may be needed for the visualization of some images.
	SW documentation and other information needed	In some cases, the supplementary notes may be needed for better understanding of the data.
	Repository for deposit of data, metadata, documentation and code	N/A
	Access restrictions	The availability of the data is according to the rule and the provisions of the signed G.A. and C.A., especially to the IPR clauses indicated there.
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A
	Standard vocabulary or mapping to commonly used ontologies	N/A
	Information Portability and exporting	N/A
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)	**Data licensing for wide reuse**	N/A
	Timing of data availability for re-use (incl. indications on embargo)	Data can be made available after publication in scientific papers.
	Data usability by Third Parties (after the end of the project)	Except for the case where results have been published, data could be reused only with the permission of UJ and according to the rules established in the G.A. and C.A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 14 – Replication study – WP9 – PARTNER: UJ		
	Restrictions to data re-use	Restrictions to the re-use of data from the Consortium Partners as imposed by the signed G.A. and C.A. rules, specific IPR established agreements and/or possible commercial confide.
	Quality assurance process	In case the characterization is made in lab, data quality is assured through the calibration of characterization equipment, and use of standards in certain measurements.
	Length of time of data re-usability	N/A
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	N/A
	Data Management Responsibilities	The UJ team is responsible for the management of the data obtained for replication study.
	Long Term Preservation assessment	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

4.1.15DATA SET n. 15 – Experimental – WP2,3,4,6,8 – PARTNER: REDSTACK

DATA SET n. 15 – Experimental– WP2,3,4,6,8 – PARTNER: REDSTACK		
1 DATA SUMMARY	**Purpose of the Data**	Data concerns physicochemical analyses of relevant water streams for the case study Number 2 (as indicated and supplied by ACSA Sorigué), optimization of ion exchange membrane preparation parameters for reverse electrodialysis, ion exchange membranes characterization, stack design development and performance evaluation studies (relative to process conditions, technologies combinations, stack configurations and setups) at the laboratory and pilot scale, process design drawing(s). All the generated data are in close relation with the implementation of the project's deliverables and objectives.
	Type and Format of data	▪ Text: field or laboratory notes, survey responses – in plain text, (txt), HTML, XLM, PDF/A {performance evaluation tests and material preparation}, ▪ Numeric: tables, counts, measurements – in .XLSX, .CSV (performance evaluation tests), ▪ Audiovisual: images, video – in .JPEG, .JPG, .PNG, .TIFF, AIFF, WAVE, .MP3, .MP4 (microscopic techniques SEM, AFM) , ▪ Models, computational models, computer code… - (for example, Python code (e.g. Py), Matlab code/scripts (e.g. M, MAT), Excel (e.g. XLSX, XLSM), ▪ Instrument specific: equipment outputs (specify equipment and format e.g. IDF) physicochemical characterization and analysis - can be converted in XLS, ASCII and CSV format}.
	Reused-Data	The majority of the data will be new, generated during the implementation of the project.
	Data origin	▪ Observational – data captured in real time (performance evaluation tests, material preparation), usually irreplaceable - i.e. sensor readings, images, telemetries, sample data. The data are reproducible if the same conditions and equipment are used. ▪ Experimental – data from lab equipment (physicochemical analysis, materials characterization), often reproducible - i.e. chromatograms, absorbance spectra, x-ray diffraction data, porometer, porosimeter. ▪ Simulation– data generated by test models where model and metadata are more important than output data. ▪ Derived/Compiled – data coming from analyses or compilation, reproducible but with high costs - i.e. text and data mining, compiled databases, 3D models. The observational and the experimental data are subject to analysis. In some case the analysis is very time consuming which increases the cost.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 15 – Experimental– WP2,3,4,6,8 – PARTNER: REDSTACK		
	Data size	Physicochemical analyses: - Revisable, - Data file size depending on the analytical technique and instrument. The Datasets vary from some Kbytes (AAS) to few Mbytes (Ion chromatography) per experiment. Material characterization: - Revisable, - Collection of data files from various laboratory instruments like SEM, AFM, contact angle, surface tension etc. Data files size depends on the instrument and (regarding microscopies) from the resolution of the images. Estimated file sizes from some KB to few MB (images) per measurement. Performance evaluation tests: - Revisable, - Files of raw data acquired by the SCADA systems of experimental devises, controllers, online sensors and analyzers etc. ASCII files (in csv format) of approximately 10 MB per study depending on testing duration and sampling rate with an overall estimated size 50 GB . - The performance evaluation provides data that can be amended. In this case the test should be repeated in order to obtain new data.
	Data Security and Storage	Supplementary and raw data initially handwritten as text in lab books are not digitized but the information is transferred to digital reports stored in office computers. Experimental data concerning physicochemical analyses of water samples or characterization of materials (e.g. membranes) are stored in the hard drives of the corresponding instrument(s). Raw data collected during performance evaluation tests by SCADA systems of experimental devices are stored as .csv or spreadsheet (excel) files in lab computers. Reports, documents, photos, designs, code, data and other digitized material generated during – and relative to - intelWATT project are stored on REDstack's computers and server. According to the lab policy all data are backed up regularly and the backups are stored in an internal network drive as well as removable hard drives. All the computers are protected by hard passwords, firewalls and antivirus / antimalware software. The current protocols prevent access to any external user.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 15 – Experimental– WP2,3,4,6,8 – PARTNER: REDSTACK		
	Data utility	The data are useful for the whole team of REDSTACK, to all members of the intelWATT project as well as to other research groups focusing on membrane science/technology and relevant applications. End users in the electric power generation and renewable energy fields, water and wastewater treatment professionals, process engineers and post graduate students focusing on the activities of intelWATT are indicative additional groups which benefit from the project's results. The data could be further exploited as source for new funding or new internal scientific projects as well as for commercialization purposes of REDstack.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 15 – Experimental– WP2,3,4,6,8 – PARTNER: REDSTACK		
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Discoverability of the data is achieved by making the analyzed data available (provided that rules and provisions of signed GA, CA and IPR interests are not violated) for publication in scientific journals. Discoverability is further maximized linking the article with the relevant dataset (including related metadata) which is deposited in a scientific data repository. These data could be further exploited as source for new funding or new internal scientific projects. The relative metadata are stored mainly in .txt, and .pdf format. Experimental keywords will be used to create the corresponding metadata. These will include some common fields like the name/code of the material (membrane), partner's acronym, as well as some specific info depending on the technique. For example, for membrane evaluation the following fields will be included: Type of experiment, Feed composition, Temperature, Pressure, Flow rates, Sample info, etc. Even though the metadata file is not created by a standard schema and is in the form of a simple .txt file, it should provide the following information: Title - Name of the dataset or research project that produced it Creator - Names and addresses of the organization or people who created the data Identifier - Number used to identify the data Subject - Keywords or phrases describing the subject or content of the data Access information - Where and how your data can be accessed by other researchers Date - Key dates associated with the data, including time period covered by the data Methodology - How the data was generated, including equipment or software used, experimental protocol, other things one might include in a lab notebook Data processing - Information on how the data has been altered or processed Sources - Citations to material for data derived from other sources, including details of where the source data is held and how it was accessed List of file names - List of all data files associated with the dataset, with their names and file extensions File Formats - Format(s) of the data, e.g. FITS, SPSS, HTML, JPEG, and any software required to read the data File structure - Organization of the data file(s) and the layout of the variables, when applicable Variable list - List and description of variables in the data files, when applicable Code lists - Explanation of codes or abbreviations used in either the file names or the variables in the data files (e.g. '999 indicates a missing value in the data')

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 15 – Experimental– WP2,3,4,6,8 – PARTNER: REDSTACK		
	Identifiability of data (refer to standard id mechanisms)	Data is made identifiable by manually naming data files according to the rules of the lab and the file naming conventions referred below. No standard ID mechanism is used. Data (in the form of document, spreadsheets, images, presentation etc.) uploaded to an online data repository is identified by a Digital Object Identification (DOI) which assigned automatically by the repository. The DOI also contains metadata that provides the relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored
	Naming conventions used	All the results and data generated during the implementation period of the project will be stored in dedicated folders. The folders' structure is tree like. The root folder (named with the project acronym "intelWATT") contains the WP folders. Further structuring inside the WP folders will be based on Task or Case study numbering. For example folder ..\\intelWATT\WP_2\TASK_3 contains files and data relative to the task 3 of WP 2. Data files are manually named in a way that their source can be uniquely distinguished. For example, a name consisting of a prefix (e.g. characteristic of the experimental technique or material) followed by a date and/or time stamp could be an appropriate convention for a file containing raw data from an experimental device's SCADA systems. Otherwise, the authors will name them as they see fit. These data are used for the preparation of internal reports. The reports are named by the project and the period, in which the data belongs. In addition, the raw data files should be linked with relevant records (in physical form in a lab book or in a text file) containing all the necessary details for the reusability / interoperability of the data (e.g. the experimental set up and conditions, the preparation method of the materials, pre and / or post treatments of the samples etc.).
	Search keywords approach	In some cases, some keywords are used including the name of the sample and the corresponding technique used.
	Clear versioning approach	Revised datasets will take the extension "rev" followed by the number of the revision made.
	Standards or procedures for metadata creation applied	N/A
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 15 – Experimental– WP2,3,4,6,8 – PARTNER: REDSTACK		
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	The data obtained by REDSTACK belongs to REDSTACK. The REDSTACK will define the way and the time that these data can be presented. The analyzed data that are presented to the consortium are available for the whole consortium for the full duration of the project and according to the rules and provisions of the signed G.A., C.A. and intellectual property obligations. Accessibility of data out of the scope of intelWATT project always requires the permission of REDSTACK. The same occurs also for the raw data. In such a case the raw data requires the permission of REDSTACK and can be made available only upon request and under separate agreement between the partners and as long this request does not interfere with future publication and intellectual property interests or precede data analysis.
	How data will be made available	Access to REDSTACK's data is restricted and always requires the permission of REDSTACK and a separate written agreement between REDSTACK and relevant partner(s), a.o. because of IP and commercial confidentiality reasons. Non-confidential data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Specific data can be made available upon requests (see 2.4).
	Methods or SW tools for data access	Tools for accessing the data are any text editor and spreadsheet packages for the output files. Optionally photo/image viewers / editors or CAD software may be needed for the visualization of some engineering designs.
	SW documentation and other information needed	In some cases, the supplementary notes may be needed for better understanding of the data.
	Repository for deposit of data, metadata, documentation and code	N/A
	Access restrictions	As a rule, the availability of the data is according to the rule and the provisions of the signed G.A. and C.A., especially to the IPR clauses indicated there. In some circumstances, REDSTACK will not be able to disclose some of this data until a certain time within the project timeframe due to possible commercial confidentiality and intellectual property reasons.
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 15 – Experimental– WP2,3,4,6,8 – PARTNER: REDSTACK		
	Standard vocabulary or mapping to commonly used ontologies	N/A
	Information Portability and exporting	N/A
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)	**Data licensing for wide reuse**	N/A
	Timing of data availability for re-use (incl. indications on embargo)	Data could be made available after its publication. Upon request to REDstack. The raw data will be not available to other partners, but can be shared only with the permission of REDSTACK.
	Data usability by Third Parties (after the end of the project)	Except for the case where results have been published, data could be reused only with the permission of REDSTACK and according to the rules established in the G.A. and C.A.
	Restrictions to data re-use	Restrictions to the re-use of data from the Consortium Partners as imposed by the signed G.A. and C.A. rules, specific IPR established agreements and/or possible commercial confidentialities
	Quality assurance process	Data quality is assured through: - Regular checking and maintenance of the experimental and mechanical equipment including electronic sensors and control devices, - Calibration of sensors (temperature, pressure, flow, conductivity etc.), - Calibration and performance verification of analytical instruments, - Use of standards for certain measurements, - Blank experiments and repetition of experiments, - Data entry validation and peer review of data, - Good practices from the researchers
	Length of time of data re-usability	N/A
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	N/A
	Data Management Responsibilities	REDSTACK is responsible for the management of their own data obtained within IntelWATT project. REDSTACK's researchers are highly educated and qualified ensuring that datasets are accurate and reliable.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 15 – Experimental– WP2,3,4,6,8 – PARTNER: REDSTACK		
	Long Term Preservation assessment	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

4.1.16DATA SET n. 16 – Experimental – WP3,6 – PARTNER: CIEMAT

DATA SET n. 16 – Experimental – WP3,6 – PARTNER: CIEMAT		
1 DATA SUMMARY	**Purpose of the Data**	Data concerns membrane module performance evaluation studies (relative to process conditions, technologies combinations, modules configurations and setups, membrane shape and dimension etc.) at the lab and pilot scale, process design drawing(s). All the generated data are in close relation with the implementation of the project's deliverables and objectives.
	Type and Format of data	▪ Text: field or laboratory notes – in plain text, (txt), HTML, XLM, PDF/A {performance evaluation tests}, ▪ Numeric: tables, counts, measurements – in .XLSX, .CSV (performance evaluation tests), ▪ Audiovisual: images – in .JPEG, .JPG, .PNG, .TIFF.
	Reused-Data	The majority of the data will be new, generated during the implementation of the project.
	Data origin	▪ Observational – data captured in real time, usually irreplaceable - i.e. sensor readings, images, sample data. The data are reproducible if the same conditions and equipment are used. ▪ Experimental – data from lab and pilot equipment (performance evaluation tests), often reproducible, but with higher costs. The observational and the experimental data are subject to analysis. In some case the analysis is very time consuming which increases the cost.
	Data size	Performance evaluation tests: - Revisable, - Files of raw data acquired by the SCADA systems of experimental devises, controllers, online sensors and analyzers etc. ASCII files (in csv format) of approximately 10 MB per study depending on testing duration and sampling rate with an overall estimated size 1 GB . - The performance evaluation provides data that can be amended. In this case the test should be repeated in order to obtain new data.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 16 – Experimental – WP3,6 – PARTNER: CIEMAT		
	Data Security and Storage	Supplementary and raw data initially handwritten as text in lab books are not digitized but the information is transferred to digital reports stored in office computers. Raw data collected during performance evaluation tests by SCADA systems of experimental devices are stored as .csv or spreadsheet (excel) files in lab computers. Reports, documents, photos, designs and other digitized material generated during – and relative to - intelWATT project are stored in office computers as well as in the professional Dropbox account of the team and EmDESK platform. According to the lab policy all data are backed up regularly and the backups are stored in an internal network drive as well as removable hard drives. All the computers are protected by hard passwords, firewalls and antivirus / antimalware software. The current protocols prevent access to any external user.
	Data utility	The data are useful for the whole team of CIEMAT, to all members of the intelWATT project as well as to other research groups focusing in membrane science/technology and relevant applications. End users in the electric power generation and renewable energy fields, water and wastewater treatment professionals, process engineers and post graduate students focusing in the activities of intelWATT are indicative additional groups which benefit from the project's results. The data can be further exploited as source for new funding or new internal scientific projects as well as for commercialization purposes.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 16 – Experimental – WP3,6 – PARTNER: CIEMAT		
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Discoverability of the data is achieved by making the analyzed data available (provided that rules and provisions of signed GA, CA and IPR interests are not violated) for publication in scientific journals. Discoverability is further maximized linking the article with the relevant dataset (including related metadata) which is deposited in a scientific data repository. These data can be further exploited as source for new funding or new internal scientific projects. The relative metadata are stored mainly in .txt, and .pdf format. Experimental keywords will be used to create the corresponding metadata. These will include some common fields like the name/code of the material (membrane or IX resin), partner's acronym, as well as some specific info depending on the technique. For example, for membrane evaluation the following fields will be included: Type of experiment, Feed composition, Temperature, Pressure, Flow rates, Energy efficiency, Sample info, etc. Even though the metadata file is not created by a standard schema and is in the form of a simple .txt file, it should provide the following information: Title - Name of the dataset or research project that produced it Creator - Names and addresses of the organization or people who created the data Identifier - Number used to identify the data Subject - Keywords or phrases describing the subject or content of the data Access information - Where and how your data can be accessed by other researchers Date - Key dates associated with the data, including time period covered by the data Methodology - How the data was generated, including equipment or software used, experimental protocol, other things one might include in a lab notebook Data processing - Information on how the data has been altered or processed

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 16 – Experimental – WP3,6 – PARTNER: CIEMAT		
		Sources - Citations to material for data derived from other sources, including details of where the source data is held and how it was accessed List of file names - List of all data files associated with the dataset, with their names and file extensions File Formats - Format(s) of the data, e.g. FITS, SPSS, HTML, JPEG, and any software required to read the data File structure - Organization of the data file(s) and the layout of the variables, when applicable Variable list - List and description of variables in the data files, when applicable Code lists - Explanation of codes or abbreviations used in either the file names or the variables in the data files (e.g. '999 indicates a missing value in the data')
	Identifiability of data (refer to standard id mechanisms)	Data is made identifiable by manually naming data files according to the rules of the lab and the file naming conventions referred below. No standard ID mechanism is used. Data (in the form of document, spreadsheets, images, presentation etc.) uploaded to an online data repository is identified by a Digital Object Identification (DOI) which assigned automatically by the repository. The DOI also contains metadata that provides the relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 16 – Experimental – WP3,6 – PARTNER: CIEMAT		
	Naming conventions used	All the results and data generated during the implementation period of the project will be stored in dedicated folders. The folders' structure is tree like. The root folder (named with the project acronym "intelWATT") contains the WP folders. Further structuring inside the WP folders will be based on Task or Case study numbering. For example folder ..\\intelWATT\WP_2\TASK_3 contains files and data relative to the task 3 of WP 2. Data files are manually named in a way that their source can be uniquely distinguished. For example, a name consisting of a prefix (e.g. characteristic of the experimental technique or material) followed by a date and/or time stamp could be an appropriate convention for a file containing raw data from an experimental device's SCADA systems. In addition, the raw data files should be linked with relevant records (in physical form in a lab book or in a text file) containing all the necessary details for the reusability / interoperability of the data (e.g. the experimental set up and conditions, the preparation method of the materials, pre and / or post treatments of the samples etc.). <u>*Second option*</u> *Data are manually named in a way that one can distinguish their source. Also the name of the interpreter/analyser is added.* *These data are used for the preparation of internal reports. The reports are named by the project and the period, in which the data belongs.*
	Search keywords approach	In some cases some keywords are used including the name of the sample and the corresponding technique used. <u>*Second option*</u> *No specific approach is provided. The pattern for keywords generation is mainly based on the kind of experiments performed, the experimental procedures and conditions, characteristics and properties of the under testing materials, analytical methods etc. As an example, for the performance evaluation of a PVDF membrane the recommended keywords are: PVDF membranes, flat sheet modules, membrane distillation, hydrophobicity, heat recovery etc.*
	Clear versioning approach	Revised datasets will take the extension "rev" followed by the number of the revision made. <u>*Second option*</u> *The versioning of data files is manual indicating the date of its creation and the writer.*

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 16 – Experimental – WP3,6 – PARTNER: CIEMAT		
	Standards or procedures for metadata creation applied	Please select between N/A and the text below Specific templates will be prepared for the collection of all the necessary information related to a particular dataset. A supplementary .pdf or .txt file will be created using the above information and a standardized layout / format. This file will be uploaded in a data repository where a DOI will be automatically assigned to it. The DOI contains metadata that provides users with relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored. Even though the project discipline is not related directly to any metadata standard or schema, however there are some general purpose standards like the Dublin Core Metadata Initiative that can be applied. Metadata Editors (preferably open access) will be used to generate partly or complete the relevant metadata files
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	N/A
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	The raw data obtained by the CIEMAT team belongs to CIEMAT. The research team will define the way and the time that these data can be presented. The analyzed data that are presented to the consortium are available for the whole consortium for the full duration of the project and according to the rules and provisions of the signed G.A., C.A. and intellectual property obligations. Accessibility of data out of the scope of intelWATT project requires the permission of CIEMAT. The same occurs also for the raw data. In such a case the raw data can be made available only upon request and under separate agreement between the partners and as long this request does not interfere with future publication and intellectual property interests, or precede data analysis.
	How data will be made available	Data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Disclosable data will be deposited in an open data repository, such as Zenodo or other recommended repositories (see below) provided that rules and provisions in signed GA and CA, IPR obligations or/and confidentiality restrictions are not violated. An option being considered is to link the data and metadata to forthcoming publications. Specific data can be made available upon requests (see 2.4).
	Methods or SW tools for data access	Tools for accessing the data are any text editor and spreadsheet packages for the output files. Optionally photo/image viewers / editors or CAD software may be needed for the visualization of some engineering designs.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 16 – Experimental – WP3,6 – PARTNER: CIEMAT		
	SW documentation and other information needed	In some cases, the supplementary notes may be needed for better understanding of the data.
	Repository for deposit of data, metadata, documentation and code	Specific repositories that will be considered include: FigShare, Mendeley Data, Open Science Framework, Harvard Dataverse, Dryad Digital Repository, Science Data Bank, Zenodo, which can be used for storage of data and metadata. These generalist repositories may also be appropriate for archiving associated analyses, or experimental-control data. *Second Option* *N/A*
	Access restrictions	As a rule, the availability of the data is according to the rule and the provisions of the signed G.A. and C.A., especially to the IPR clauses indicated there. In some circumstances, NCSR will not be able to disclose some of this data until a certain time within the project timeframe due to possible commercial confidentiality and intellectual property reasons.
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A
	Standard vocabulary or mapping to commonly used ontologies	N/A
	Information Portability and exporting	N/A
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)	**Data licensing for wide reuse**	N/A
	Timing of data availability for re-use (incl. indications on embargo)	Data can be made available after publication in scientific papers. The raw data will be not available to other partners, but they can be shared only after permission of the CIEMAT team.
	Data usability by Third Parties (after the end of the project)	Except for the case where results have been published, data could be reused only with the permission of CIEMAT and according to the rules established in the G.A. and C.A.
	Restrictions to data re-use	Restrictions to the re-use of data from the Consortium Partners as imposed by the signed G.A. and C.A. rules, specific IPR established agreements and/or possible commercial confidentialities.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 16 – Experimental – WP3,6 – PARTNER: CIEMAT		
	Quality assurance process	Data quality is assured through the: - Regular checking and maintenance of the experimental and mechanical equipment including electronic sensors and control devices, - Calibration of sensors (temperature, pressure, flow, conductivity etc.), - Calibration and performance verification of analytical instruments, - Use of standards for certain measurements, - Blank experiments and repetition of experiments, - Data entry validation and peer review of data.
	Length of time of data re-usability	N/A
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	N/A
	Data Management Responsibilities	The CIEMAT team is responsible for the management of the data obtained within the frame of intelWATT project. Researchers of CIEMAT are high educated and qualified to ensure that dataset is accurate and reliable.
	Long Term Preservation assessment	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

4.1.17DATA SET n. 17 – Experimental – WP1,2,4,5,6,7,9,10 – PARTNER: NOKIA

DATA SET n. 17 – Experimental– WP1,2,4,5,6,7,9,10 – PARTNER: NOKIA		
1 DATA SUMMARY	**Purpose of the Data**	Data concerns physicochemical analyses of relevant water streams for the case study No1 (as indicated and supplied by Partner No4, Public Power Company, Greece), and in particular Contaminants/species to be detected, as well as water parameters to be monitored; Also data will be measured with sensors and actuators developed by Partner No11 Fuelics and Partner No14 Techedge. Measurement data are not data produced by NOKIA, only propagated by NOKIA. Sensors will be connected to NOKIA's IoT platform. Via the IoT platform (sensors) data will be conveyed to the data Lake and analysis applications developed by Partner No8 POLITO. All the generated data are in close relation with the implementation of the project's deliverables and objectives. Operation and maintenance data, generated by software, used as a mediation layer between IMPACT IoT platform and the sensors is considered as data produced by NOKIA. The same applies for any data related to the operation and maintenance of the IoT platform.
	Type and Format of data	▪ Text: field or laboratory notes, survey responses – in plain text, (txt), HTML, XLM, PDF/A {performance evaluation tests and material preparation}, ▪ Numeric: tables, counts, measurements – in .XLSX, .CSV (performance evaluation tests), ▪ Audiovisual: images, video – in .JPEG, .JPG, .PNG, .TIFF, AIFF, WAVE, .MP3, .MP4 (microscopic techniques SEM, AFM), ▪ Instrument specific: sensor outputs {physicochemical characterization and analysis - can be converted in XLS, ASCII and CSV format}, ▪ Computer code in Golang or any other suitable programming language, developed to implement the mediation layer between sensors and/or sensors interface and IMPACT IoT platform, ▪ Messages content and format in JSON, XML HTTP, or any other proprietary / non proprietary protocol, exchanged between sensors and IMPACT IoT platform. Also messages content and format in JSON, XML HTTP, or any other proprietary protocol, exchanged between IMPACT IoT platform and the Data Lake, as well as between IMPACT IoT platform and applications.
	Reused-Data	The majority of the data will be new, generated during the implementation of the project.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 17 – Experimental– WP1,2,4,5,6,7,9,10 – PARTNER: NOKIA		
	Data origin	▪ Observational – data captured in real time (performance evaluation tests, material preparation, software tracers and logs), usually irreplaceable - i.e. sensor readings, sample data. The data are reproducible if the same conditions and equipment are used. ▪ Experimental – data from lab equipment, often reproducible. The observational and the experimental data are subject to analysis. In some case the analysis is very time consuming which increases the cost.
	Data size	Data size depending on data produced by partners sensors, as described above, typically: - Revisable, - Collection of data files from various laboratories. Estimated file sizes from some KB to few MB (images) per measurement.
	Data Security and Storage	Supplementary and raw data initially handwritten as text in lab books are not digitized but the information is transferred to digital reports stored in office computers. Experimental data concerning measurements, from partners' sensors, which are conveyed to the IMPACT server. After their transmission IMPACT platform deletes any data. Computer programs, pieces of code (functional or non-functional) Reports, documents, photos, designs and other digitized material generated during – and relative to - intelWATT project are stored in office computers as well as in cloud resources (OneDrive resources used by Nokia). All the computers are protected by hard passwords, firewalls and antivirus / antimalware software. The current protocols prevent access to any external user.
	Data utility	The data is useful for the whole team of Nokia Hellas, to all members of the intelWATT project as well as to other research groups focusing in membrane science/technology and relevant applications. The data could be also useful to IoT related enterprises and/or research institutes. End users in the electric power generation and renewable energy fields, water and wastewater treatment professionals, process engineers and post graduate students focusing in the activities of intelWATT are indicative additional groups which benefit from the project's results. The data can be further exploited as source for new funding or new internal scientific projects as well as for commercialization purposes and Research and Development projects.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 17 – Experimental– WP1,2,4,5,6,7,9,10 – PARTNER: NOKIA		
2.1 FAIR DATA - Making data findable	**Discoverability of data (metadata provision)**	Discoverability of the data is achieved by making the analyzed data available (provided that rules and provisions of signed GA, CA and IPR interests are not violated) for publication in scientific journals. Operation and maintenance data generated by software, used as a mediation layer between IMPACT IoT platform and the sensors will be discoverable, using common software engineering techniques. Discoverability is further maximized linking the article with the relevant dataset (including related metadata) which is deposited in a scientific data repository. These data can be further exploited as source for new funding or new internal scientific projects. The relative metadata are stored mainly in .txt, and .pdf format. Experimental keywords will be used to create the corresponding metadata. These will include some common fields like the name/code of the measured entity (chemical or other), partner's acronym, as well as some specific info depending on the technique. For example: Type of experiment, Feed composition, Temperature, Pressure, Flow rates, Sample info, etc. Even though the metadata file is not created by a standard schema and is in the form of a simple .txt file, it should provide the following information: Title - Name of the dataset or research project that produced it Creator - Names and addresses of the organization or people who created the data Identifier - Number used to identify the data Subject - Keywords or phrases describing the subject or content of the data Access information - Where and how your data can be accessed by other researchers Date - Key dates associated with the data, including time period covered by the data Methodology - How the data was generated, including equipment or software used, experimental protocol, other things one might include in a lab notebook Data processing - Information on how the data has been altered or processed

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 17 – Experimental– WP1,2,4,5,6,7,9,10 – PARTNER: NOKIA		
		Sources - Citations to material for data derived from other sources, including details of where the source data is held and how it was accessed List of file names - List of all data files associated with the dataset, with their names and file extensions File Formats - Format(s) of the data, e.g. FITS, HTML, JPEG, and any software required to read the data File structure - Organization of the data file(s) and the layout of the variables, when applicable Variable list - List and description of variables in the data files, when applicable Code lists - Explanation of codes or abbreviations used in either the file names or the variables in the data files (e.g. '999 indicates a missing value in the data')
	Identifiability of data (refer to standard id mechanisms)	Data is made identifiable by manually naming data files according to the rules of the lab and the file naming conventions referred below. No standard ID mechanism is used. Data (in the form of document, spreadsheets, images, presentation, computer code etc.) uploaded to an online data repository is identified by a Digital Object Identification (DOI) which assigned automatically by the repository. The DOI also contains metadata that provides the relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored.
	Naming conventions used	All the results and data generated during the implementation period of the project will be stored in dedicated folders. The folders' structure is tree like. The root folder (named with the project acronym "intelWATT") contains the WP folders. Further structuring inside the WP folders will be based on Task or Case study numbering. For example folder ..\\intelWATT\WP_2\TASK_3 contains files and data relative to the task 3 of WP 2. Data files are manually named in a way that their source can be uniquely distinguished. In addition, the raw data files should be linked with relevant records (in physical form in a lab book or in a text file) containing all the necessary details for the reusability / interoperability of the data (e.g. the experimental set up and conditions, the preparation method of the materials, pre and/or post treatments of the samples etc.). *Second option* *Data are manually named in a way that one can distinguish their source. Also the name of the interpreter/analyser is added.* *These data are used for the preparation of internal reports. The reports are named by the project and the period, in which the data belongs.*

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 17 – Experimental– WP1,2,4,5,6,7,9,10 – PARTNER: NOKIA		
	Search keywords approach	In some cases some keywords are used including the name of the sample and the corresponding technique used. <u>*Second option*</u> *No specific approach is provided. The pattern for keywords generation is mainly based on the kind of experiments performed, the experimental procedures and conditions, characteristics and properties of the under testing materials, analytical methods etc. As an example, for the performance evaluation of a PVDF membrane the recommended keywords are: PVDF membranes, flat sheet modules, membrane distillation, hydrophobicity, heat recovery etc.*
	Clear versioning approach	Revised datasets will take the extension "rev" followed by the number of the revision made. <u>*Second option*</u> *The versioning of data files is manual indicating the date of its creation and the writer.*
	Standards or procedures for metadata creation applied	Please select between N/A and the text below Specific templates will be prepared for the collection of all the necessary information related to a particular dataset. A supplementary .pdf or .txt file will be created using the above information and a standardized layout / format. This file will be uploaded in a data repository where a DOI will be automatically assigned to it. The DOI contains metadata that provides users with relevant information about an object, such as the title, author, keywords, year of publication and the URL where that document is stored. Even though the project discipline is not related directly to any metadata standard or schema, however there are some general purpose standards like the Dublin Core Metadata Initiative that can be applied. Metadata Editors (preferably open access) will be used to generate partly or complete the relevant metadata files
2.2. DATA GROUPING and DISTRIBUTION	**Data Grouping in record**	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 17 – Experimental– WP1,2,4,5,6,7,9,10 – PARTNER: NOKIA		
2.3 FAIR DATA – Making data openly accessible	**Data openly available or kept close**	The raw data (i.e. the operation and maintenance data generated by software, used as a mediation layer between IMPACT IoT platform and the sensors) obtained by NOKIA team belongs to NOKIA. The research team will define the way and the time that these data can be presented. The analyzed data that are presented to the consortium are available for the whole consortium for the full duration of the project and according to the rules and provisions of the signed G.A., C.A. and intellectual property obligations. Accessibility of data out of the scope of intelWATT project requires the permission of NOKIA. The same occurs also for the raw data. In such a case the raw data can be made available only upon request and under separate agreement between the partners and as long this request does not interfere with future publication and intellectual property interests, or precede data analysis.
	How data will be made available	Data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Data can be made available to the Consortium through the Collaborative (EMDESK) platform or website of the project. Disclosable data will be deposited in an open data repository that will be decided, given that rules and provisions in signed GA and CA, IPR obligations or/and confidentiality restrictions are not violated. An option being considered is to link the data and metadata to forthcoming publications. Specific data can be made available upon requests (see 2.4).
	Methods or SW tools for data access	Tools for accessing the data are any text editor and spreadsheet packages for the output files. Optionally photo/image viewers / editors may be needed for the visualization of some engineering designs.
	SW documentation and other information needed	In some cases, the supplementary notes may be needed for better understanding of the data.
	Repository for deposit of data, metadata, documentation and code	N/A
	Access restrictions	As a rule, the availability of the data is according to the rule and the provisions of the signed G.A. and C.A., especially to the IPR clauses indicated there. In some circumstances, NOKIA will not be able to disclose some of this data until a certain time within the project timeframe due to possible commercial confidentiality and intellectual property reasons.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

DATA SET n. 17 – Experimental– WP1,2,4,5,6,7,9,10 – PARTNER: NOKIA		
2.3 FAIR DATA – Making data interoperable	**Data interoperability assessment**	N/A
	Standard vocabulary or mapping to commonly used ontologies	N/A
	Information Portability and exporting	N/A
2.4 FAIR DATA – Increase data re-use (through clarifying licenses)	**Data licensing for wide reuse**	N/A
	Timing of data availability for re-use (incl. indications on embargo)	Data can be made available after publication in scientific and/or white papers. The raw data (i.e. the operation and maintenance data generated by software, used as a mediation layer between IMPACT IoT platform and the sensors) will be not available to other partners but can be shared only after permission of NOKIA.
	Data usability by Third Parties (after the end of the project)	Except for the case where results have been published, data could be reused only with the permission of NOKIA and according to the rules established in the G.A. and C.A.
	Restrictions to data re-use	Restrictions to the re-use of data from the Consortium Partners as imposed by the signed G.A. and C.A. rules, specific IPR established agreements and/or possible commercial confidentialities.
	Quality assurance process	Data quality is assured through the continuous maintenance of the hardware infrastructure that is dedicated to the project, and the software platforms and addons that will be needed to smoothly integrate with sensors. The maintenance task includes regular health, performance etc. tests, as well as individual quality testing of the new software addons to ensure that expectations are met.
	Length of time of data re-usability	N/A
3 ALLOCATION OF RESOURCES	**Costs estimates for making data FAIR**	N/A
	Data Management Responsibilities	NOKIA research team is responsible for the management of the data obtained within the frame of IntelWATT project. NOKIA's internal processes guarantee the quality of the data and research team is qualified to ensure that dataset is accurate and reliable.
	Long Term Preservation assessment	N/A

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

5 Open Access to Publications

The approach hereafter indicated is derived from:

- the *H2020 Programme - Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020* (Version 3.1 of 25 August 2016)
- the H2020 Online Manual section dedicated to Open Access available at http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm,

considering the terms of article 29.2 of the Grant Agreement.

Article 29.2 of the Model Grant Agreement define that under Horizon 2020 **each beneficiary must ensure open access to all peer-reviewed scientific publications relating to the Project results**. Specifically:

> ***29.2 Open access to scientific publications***
>
> *Each beneficiary must ensure open access (free of charge online access for any user) to all peer-reviewed scientific publications relating to its results.*
>
> *In particular, it must:*
>
> *(a) as soon as possible and at the latest on publication, deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications;*
>
> *Moreover, the beneficiary must aim to deposit at the same time the research data needed to validate the results presented in the deposited scientific publications.*
>
> *(b) ensure open access to the deposited publication — via the repository — at the latest:*
>
> *(i) on publication, if an electronic version is available for free via the publisher, or*
>
> *(ii) within six months of publication (twelve months for publications in the social sciences*
>
> *and humanities) in any other case.*
>
> *(c) ensure open access — via the repository — to the bibliographic metadata that identify the deposited publication.*
>
> *The bibliographic metadata must be in a standard format and must include all of the following:*
>
> *- the terms "European Union (EU)" and "Horizon 2020";*
>
> *- the name of the action, acronym and grant number;*
>
> *- the publication date, and length of embargo period if applicable, and*
>
> *- a persistent identifier.*

Therefore, partners will at least ensure that scientific peer-reviewed publications related to the project results can be read online, downloaded and printed.

Moreover, they will consider where possible to provide further options that make publications more useful for re-use, such as the right to copy, distribute, search, link, crawl and mine.

Partners are also encouraged to provide open access to monographs, books, conference proceedings and grey literature (e.g. relevant reports).

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

5.1 Granting Open Access procedure

The procedure to grant open access includes in any case two (interrelated) steps:

1. depositing publications in repositories
2. providing open access to them and the related bibliographic metadata

STEP 1 - DEPOSITING PUBLICATIONS IN REPOSITORIES

Partners will deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications.

This will be done as soon as possible and at the latest upon publication. Where possible, the version deposited will be identical to the published version (in layout, pagination, etc.).

This step will **apply even where open access publishing** (**'gold' open access**) **is chosen,** to ensure long term preservation.

Researchers will refer to scientific repositories as for provision of each academic entity. Otherwise, they will refer to *The Open Access Infrastructure for Research in Europe (OpenAIRE)* or other common lists (such as OpenDOAR or ROAR) to determine what repository to choose, considering that acceptable choices include institutional, subject-based and centralised repositories, and exclude repositories that claim rights over deposited publications and preclude access.

In general, the use of **Zenodo** (an OpenAIRE and CERN collaboration, allowing researchers to deposit both publications and data, while providing tools to link them) is suggested if the researcher academic entity does not provide an institutional repository.

STEP 2 - PROVIDING OPEN ACCESS TO PUBLICATIONS

To ensure open access to the peer-reviewed articles deposited on the chosen repository, one of the two options will be applied:

1. **Self-archiving / 'green' Open Access**:
   Open access is granted depositing the final peer-reviewed manuscript in an open repository of choice, within at most 6 months from the publication.
   To be considered that some publishers impose embargo periods.
2. **Open access publishing / 'gold' Open Access**:
   Open Access can be granted also directly in open access journals, or in hybrid journals that allow for both selling subscriptions and making individual articles openly accessible. Monographs can also be published with this model.
   'Article processing charges' are eligible for reimbursement during the duration of the project (as other costs defined in Article 6.2.D.3 of the Model Grant Agreement).

In order to make it easier to find publications and ensure that EU funding is acknowledged as required, the partners will also provide the **bibliographic metadata** that identify the deposited publication. These will be in a standard format and include the following:

- the terms *"European Union (EU)"* & *"Horizon 2020"*
- the name of the action, acronym and grant number *(intelWATT GA N° 958454)*
- the publication date
- the length of the embargo period (if applicable)
- a persistent identifier (i.e. Digital Object Identifier – DOI, or others).

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".

5.2 Open Access to Underlying Data

Article 29.2 of the Model Grant Agreement also mentions that beneficiaries must aim to deposit at the same time as the publication the research data needed to validate the results presented in the deposited scientific publications ('underlying data'), ideally in a data repository.

The partners consider focusing specifically on research data (being it results of experiments, measurements, models or images) that is not only available in digital form, but also at some degree standardized and verifiable.

6 Conclusions

The project approach to Data Management considering the different datasets handled in the project has been defined.

The potential of making research data findable, accessible, interoperable and reusable (FAIR) has been assessed, responding to the common aim to improve access to scientific information and increase the impact of public investments in research, but in the context of a research project with such an extremely high industrial interest, data are mainly considered confidential and will not be disclosed.

In any case, the approach to Open Access for information about project results and underlying datasets that are considered disclosable has been outlined, in line with the requirements of the Grant Agreement.

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958454".