Ref. Ares(2023)3735005 - 30/05/2023

DATA MANAGEMENT PLAN

Deliverable 1.1

WP 1 MANAGE- coordination and project management

Author: Nuria Claramunt, ICRA

Date:11/05/2023

Funded by the European Union

GRANT AGREEMENT NUMBER	101081728		
FULL TITLE / ACRONYM	INNOVATIVE TOOLS TO CONTROL ORGANIC MATTER AND DISINFECTION BYPRODUCTS IN DRINKING WATER / intoDBP		
START DATE	01/12/2022	**DURATION**	48 months
END DATE	30/11/2026		
PROJECT URL	https://intodbp.eu/		
DELIVERABLE	**D 1.1**		
WORK PACKAGE	**WP1**		
CONTRACTUAL DATE OF DELIVERY	**31.05.2023**		
ACTUAL DATE OF DELIVERY	**31.05.2023**		
NATURE	**OTHER**	**DISSEMINATION LEVEL**	PUBLIC
LEAD BENEFICIARY	**ICRA**		
RESPONSIBLE AUTHOR	NURIA CLARAMUNT (ICRA)		
CONTRIBUTIONS FROM	MARIA JOSÉ FARRÉ (ICRA)		
ABSTRACT	This document describes the data management life cycle for the data to be collected, processed and/or generated by the intoDBP project. The purpose of this document is to make the research data findable, accessible, interoperable, and reusable (FAIR). It specifies how data will be handled both during and after the research project and reflects on data collection, data storage, data security and data retrieval. The intoDBP Data Management Plan is an evolving document that will be edited and updated throughout the project.		

Disclaimer

Any dissemination of results reflects only the author's view, and the European Commission is not responsible for any use that may be made of the information it contains.

Copyright message

© intoDBP Consortium, 2022
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation, or both. Reproduction is authorised provided the source is acknowledged.

TABLE OF CONTENTS

LIST OF TABLES 6
LIST OF FIGURES 6
LIST OF ABBREVIATIONS 6
EXECUTIVE SUMMARY 7
1.0 INTRODUCTION 8
1.1 What is a DMP 8
1.2 Purpose of this document 8
1.3 Objective of the initial DMP 8
1.4 Structure of this document 9
2.0 DATA SUMMARY 9
2.1 Purpose of the data collection/generation and relation to the project objectives 9
2.2 Origin of the data 10
2.3 Reused existing data 11
2.4 Expected size of the data 12
2.5 Data utility 12
3.0 FAIR DATA 12
3.1 Making Data Findable including provisions for metadata 12
3.1.1 Naming conventions 12
3.1.2 Use of standards for metadata creation 13
3.2 Making Data openly Accessible 13
3.2.1 Open Data repositories 14
3.3 Making Data Interoperable 16
3.4 Increase Data-Reuse 16
4.0 ALLOCATION OF RESOURCES 16
4.1 Costs for making your data FAIR and long-time preservation 16
4.2 Responsibilities for data management 16
5.0 DATA SECURITY 17
5.1 Data recovery, secure storage, and transfer of sensitive data 17
6.0 ETHICAL ASPECTS 20
6.1 Introduction 20
6.2 Human Ethics 20
6.3 Personal Data Processing 20
6.3.1 Personal Data Processing in CS area of Spain 20

6.3.1 Personal Data Processing in CS area of Ireland 22
6.3.1 Personal Data Processing in CS area of Cyprus 22
7.0 OTHER 22
8.0 REFERENCES 22
9.0 APPENDICES 23
9.1 Ethic issues 23
9.2 List of datasets 24

LIST OF TABLES

Table 1. List of every type of data and its description
Table 2. List of Open Data repositories
Table 3. List of data storage and recovery of some partners

LIST OF FIGURES

Figure 1: Data life cycle

LIST OF ABBREVIATIONS

Abbreviation	Description
CS	Case Study
D	Deliverable
DBP	Disinfection ByProducts
DOI	Digital Object Identifier
DOM	Dissolved Organic Matter
DMP	Data Management Plan
EC	European Commission
FAIR	Findable, Accessible, Interoperable and Reusable
GDPR	EU General Data Protection Regulation
ISIMIP	Inter-Sectoral Impact Model Intercomparison Project
PIA	Privacy Impact Assessment
DPIA	Data Protection Impact Assessment
SAR	Subject Access Request
THM	Trihalomethanes
UV	Ultra-violet

EXECUTIVE SUMMARY

The current document aims to provide a guide on how to collect, produce, manage, and reuse data of the Project “Innovative Tools to control organic matter and disinfection byproducts in drinking water, intoDBP” Grant Agreement No 101081728 for the intoDBP consortium.

This Data Management Plan (DMP) presents a data summary, describes the provisions for sharing FAIR data, and addresses data security. It also addresses the allocation of resources and data management roles and responsibilities.

The DMP is not a fixed document and will be constantly updated throughout the life of the project. More specifically, this document is the first version of the DMP, delivered in 6 Month of the project (May 2023).

1.0 INTRODUCTION

1.1 What is a DMP

The Data Management Plan of the intoDBP project describes the complete life cycle of data generated throughout the project. The DMP describes the different types of data that will be generated and collected over the lifetime of the project, the way in which the research data and parts of the datasets will be preserved and then distributed and reused.

The Data Management Plan (DMP) is a formal requirement set up by the European Commission (EC) to those projects funded by the Horizon Europe program. The present document has been prepared with aid of the DMP Online tool of Catalan Open Research Area.

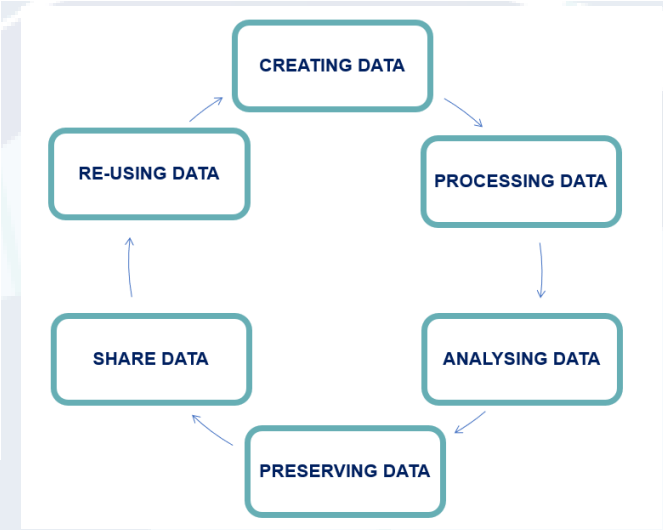


Figure 1. Data life cycle

1.2 Purpose of this document

The purpose of deliverable D1.1: Data Management Plan is to provide the key elements of the data management that will be applied in the intoDBP project and will be used by the Consortium regarding project data.

1.3 Objective of the initial DMP

The basic objective of the current document is to specify how data will be collected, produced, processed, catalogued, disseminated, and reused.

The DMP is not a fixed document but will be evolved as the project develops. This document represents the initial version of the data management life cycle for all datasets to be collected, processed, or generated by the project partners. To accomplish proper and effective management of the project's data, an overview of the datasets to be generated by the intoDBP project is presented. Subsequent versions of DMP will include

more details. The DMP reports will be updated in Month 36 (D1.2) and Month 48 (D1.3) of the project, respectively.

1.4 Structure of this document

This deliverable is structured as follows:

- Section 2 provides a data summary of the project. The DMP states the purpose of data generation and collection, explains the relation to the project objectives, specifies the types and formats of data and existing reused data as well as the data utility.
- Section 3 presents the process of making the data 'FAIR', that is making it findable, openly accessible, interoperable, and re-usable.
- Section 4 refers to the allocation of resources for the data collection and management.
- Section 5 refers to the crucial aspects of data security.
- Section 6 presents the ethical issues for the execution of intoDBP.
- Section 7 lists some of the partners' procedures for data management.
- Section 8 lists the references used to write intoDBP DMP.

2.0 DATA SUMMARY

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

The scope of the data collection and generation in intoDBP is to develop, scale-up, validate and benchmark innovative tools and strategies to protect catchments and minimize human exposure to disinfection byproducts (DBPs).

intoDBP will produce several datasets during the lifetime of the project. The nature of the data will be both quantitative and qualitative and will be analysed from a range of perspectives for project development and scientific purposes. The completion of the work plans associated to the 7 WPs of intoDBP will generate new and original scientific and technical data. Data will also be reused from previous scientific publications and will serve as reference cases, policy briefings, results, or knowledge for new research developments.

The data collection, selection, classification, and preservation are critical actions which will be maintained and carefully monitored all along the execution of the project. It will enable to exchange relevant technical information among the beneficiaries and therefore increase the efficiency of the collaborative research work for the achievement of the project.

2.2 Origin of the data

The data will be originated from several sources, which will include partner's pre-existing data, data from scientific literature, real-world measurement data, experimental data, data from simulation experiments and data from population surveys.

In intoDBP project, the following type of data will be documented:

Table 1. List of every type of data and its description.

Type of data	Description
Observational data	This type of data will be captured real-time and, typically, cannot be reproduced exactly which makes the strategies applied for documentation of this type of data extra important. Therefore, a careful and elaborated preparation of the way that data will be captured (instruments, calibration, conditions, and observation method) is crucial before recording this type of data.
Experimental data	This data will be achieved from the laboratory analysis or from equipment used. For this type of data, a thorough and elaborated preparation is needed before the capturing. Such data can be reproduced if needed, but it may be time-consuming and expensive.
Simulation data	This type of data will be obtained from models. The reproducibility of this data is subjected to knowing the input, version control of the input and software code as well as information about running environment such as operating system, release date, software dependencies, etc.
Derived or compiled data	Data derived upon transformation of existing datasets to create new data. This type of data will be obtained from data mining or statistical analysis.
Design data	Design data refers to the information and specifications used during the process of creating and developing a product, system, or project. It includes various elements such as drawings, diagrams, blueprints, models, technical specifications, and other relevant documentation that describe the intended design and functionality of the entity being created.
Metadata	Metadata are the information about the data. The reproducibility of metadata depends on the availability of the original data that metadata gives information about.
Methodology data	This data will be achieved from the standard operation procedures or technical specification of instrument/process.

Project management data	This type of data will be generated during the project by the intoDBP consortium. The file formats to be generated and collected are expected to be presentations and meeting minutes.
Dissemination data	This type of data will be obtained from all dissemination and communication activities of intoDBP project.

Data collection process

This will be provided by Work Package Leaders during the project implementation and will be documented in the 2nd and 3rd versions of this deliverable.

Formats of the data

The formats of the files to be generated and collected are expected to be:

- Graphics: jpeg, odg, pdf, png, pttx
- Tables: odsu, opj, xlsx, csv
- Text: docx, pdf, txt, JCAMP-DX
- Software: dwg, m, R, py, shp, gpkg, Postgre SQL
- Others: raw

2.3 Reused data

Although intoDBP aspires to produce its own data, it will also reuse in the most suitable way any existing data that can contribute to its purpose. In this regard, the following data can be reused:

- Historical data of UV sensors measurements of dissolved organic matter (DOM) to develop an early warning detection tool of DBPs (Task 3.4).
- Historical meteorological data and DOM data together with archived weather forecast data to develop a modelling workflow for short-term prediction changes in water quality (Task 5.2).
- ISIMIP climate data to develop long-term predictive models for assessing long-term in DOM content in source water and DBP formation (Task 5.3).
- Historical data on water use collected by on-going epidemiological studies in Europe (Task 6.1).
- Data generated at WP 2 to develop a model for each DBP to advance the knowledge on exposure to DBP mixtures (Task 6.2).
- Long time period of monitored water quality parameters and climate variables to describe temporal patterns of THM concentrations and the climate variables (Task 6.3).

2.4 Expected size of the data

The size of the data is today not known. Size is relevant for future data-sharing and making data accessible. In the course of the time, information about the size of the managed data in intoDBP project should be revised and updated accordingly.

2.5 Data utility

Within the Consortium:

The data sets will be shared within the consortium as the working baseline to produce the scientific publications, policy briefs and an intoDBP guidebook.

Beyond the Consortium:

The data can be used by independent researchers to understand better the contents and conclusions of the scientific publications, which base their findings on the data. Furthermore, independent researchers can use the files to produce figures and publications, showing comparisons of their own results and the intoDBP results. Scientists can also use the data files to repeat the experiments and measurements to verify and validate the intoDBP research. Finally, the data sets may also be used by scientific writers and the press to produce high-quality infographics.

Impacts (benefits) beyond academia:

The data may be also useful to communities of professors, students, engineers, managers, and policy makers, during and after the end of the project. By improving water quality, promoting sustainable management, and raising awareness, the project will contribute to healthier communities, environmental sustainability, and resilient water systems, thereby maximizing its overall influence on society.

3.0 FAIR DATA

3.1 Making Data Findable including provisions for metadata

Data must be findable easily, rapidly, and identically. Therefore, exact, and standard measures to identify datasets will be defined and updated in the next versions of DMP.

All open project results deposited in a repository will provide search keywords together with their metadata. Keywords for open data will be selected from controlled vocabularies that are suitable for the specific type of data.

Version control mechanisms should be established and documented before any data are collected or generated.

3.1.1 Naming conventions

Files will be structured in terms of project and lead partner and publication id and figure and filenames.

The proposed naming convention for datasets in the intoDBP project consists of three parts:

- A prefix indicating the data category (e.g. metadata, backup data, template)
- A short name for the dataset
- An integer indicating the version of the dataset

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

3.1.2 Use of standards for metadata creation

Metadata are created manually by depositors in the deposit form at the repository. Each dataset associated with data will be accompanied with unique specified metadata to allow ease of access and re-usability.

Standards for metadata creation, which addresses issues in the metadata, will be considered.

3.2 Making Data openly Accessible

In the intoDBP project, research data needed to validate the results associated with scientific publications will be made openly available as the default. Other datasets are proprietary and have high commercial sensitivity. In the cases where private data are processed, permission will be requested by the provider prior to making the data publicly available.

Three different access levels will be defined:

- Open access refers to level of access with no restriction. It means that the information, data or resource is available to anyone without any restrictions or limitations. There are not special requirements necessary to access the information.
- Restrictive access refers to a level of access control or restrictions placed on the availability and usage of the dataset. It means that access to the dataset is limited and may require certain criteria or permissions to be met before accessing the data. These restrictions can be in the form of access licenses, data sharing agreements, or other mechanisms to ensure that only authorized individuals or entities can access and use the dataset. The purpose of implementing a restrictive access level is typically to protect sensitive or confidential information, comply with legal and ethical obligations, or maintain control over the dataset's usage and distribution.
- Confidential access implies stricter restrictions. It is typically used for sensitive and classified information that requires protection from unauthorized disclosure. Confidential access is granted only to individuals or groups who have been explicitly authorized and have a legitimate need-to-know. This level often involves

additional security measures, such as encryption, logging, and monitoring, to ensure the confidentiality and integrity of the information.

Once processing, quality control, organisation, analysis, and publication are complete, the data will be made accessible by deposition in the open access repositories.

The data will be accessible using standard publicly available software where possible, but if specialist software tools are developed, i.e. created within Matlab or other software programmes, a short text file (e.g. ASCII) will be provided with the data file to explain the software required.

WP2, WP3, WP4, WP5, and WP6 are working with sensitive data. WP6 will receive personal data in questionnaires, such data must be anonymised before making available outside the project. A procedure for handling personal data is describe in section 6 of this DMP.

Users will be required to acknowledge the consortium and the source of the data in any resulting publications.

Identity of the person accessing the data will not be directly ascertained. However, we expect users to follow the standard norms of scientific citation and use of the data in this context will be tracked through scientific citation.

3.2.1 Open Data repositories

The open data repositories, which will be used during the project, are:

Table 2. List of Open Data repositories

Name	Catalan Open Research Area (CORA) Repository
Website	https://dataverse.csuc.cat/
Description	It is an open access multidisciplinary repository in FAIR mode following the European Open Science Cloud (EOSC) guidelines. It works with the open source Dataverse software, developed by Harvard's Institute for Quantitative Social Science.
Partners	Fundació Institut Català de Recerca de l'Aigua (ICRA) and Fundación Privada Instituto de Salud Global Barcelona (ISGLOBAL)
Name	Zenodo and GitHub Repository
Website	https://zenodo.org/
Description	It is the multidisciplinary open data repository in Europe commissioned by the EC. Zenodo code is itself open source. GitHub is a web-based platform used for version control and collaboration in software development projects.

Partners	KIOS Research and Innovation Centre of Excellence and NIREAS International Water Research Centre. Both research centres belong to University of Cyprus (UCY)
Name	STÓR
Website	https://eprints.dkit.ie/
Description	It is the Dundalk Institute of Technology's Open Access (OA) Institutional Repository. It works with Eprints, which is an open-source digital repository platform developed by the University of Southampton.
Partners	Dundalk Institute of Technology (DkIT)
Name	Il portale Italiano dell'Open Data (DatiOpen)
Website	http://www.datiopen.it/it
Description	It is an Italian multidisciplinary open data repository. It works with StatPortal Open Data platform.
Partners	Aquasoil SRL and Universita Degli Studi di Catania (UNICT)
Name	Cooperation Open Government Data Österreich
Website	https://www.data.gv.at/
Description	It is a multidisciplinary open data repository of the Austrian government.
Partners	S::can GmbH and Scan Iberia
Name	Hardvard Dataverse Repository
Website	https://dataverse.harvard.edu/
Description	It is a multidisciplinary open-source data repository developed by Harvard's Institute for Quantitative Social Science.
Partners	The Hebrew University of Jerusalem (HUJI)

Canal de Isabel II Gestión SA (CANAL), Ente de Abastecimiento de Agua Ter-Llobregat (ATL) and Water Board of Lemesos are companies that manage the integral water cycle and they do not use data repositories open to the general public. Any data they generate can be shared with the Consortium for analysis, but always in accordance with the confidentiality of the Consortium's agreement.

The open data repositories used by Water Europe (WE) and SK Euromarket will be updated in the next version of this DMP. However, any data they generate can be shared with the Consortium for analysis, subject to the strict confidentiality requirements stipulated in the Consortium's agreement.

3.3 Making Data Interoperable

As intoDBP project will be making use of different data sources along with the generated data by the project itself, interoperability is a major challenge to be tackled for making the provided components and datasets valuable and usable. By promoting interoperability, the plan aims to facilitate collaboration and maximize the potential value of the data. It ensures that data can be easily understood, processed, and combined with other datasets, thereby enabling comprehensive analysis and deriving meaningful insights.

Seamless integration and exchange of data across different systems and platforms will be ensured during the lifetime of the project. Interoperability will be achieved by adhering to standardized formats, protocols, and metadata structures, enabling efficient data sharing, integration, and analysis. The detailed data processing frameworks and data formats will be described in the updated versions of the DMP in Month 36 (D1.2) and Month 48 (D1.3) of the project, respectively.

3.4 Increase Data-Reuse

For our research data collection, the quality control of the data can happen at various stages during the quality assurance process. The initial quality control of the data, during data collection, is the primary responsibility of the project data creator/owner, who must ensure that the recorded data reflect the facts, responses, observations, and events. The quality of the data collection methods used strongly influences data quality, and documenting in detail how data are collected provides evidence of such quality. Errors can also occur during data entry. Data are digitised, transcribed, entered in a database or spreadsheet, or coded.

The data will remain re-usable after the end of the project by anyone interested in it. The open data repositories have time horizons of at least 10 years.

4.0 ALLOCATION OF RESOURCES

4.1 Costs for making your data FAIR and long-time preservation

The cost of data management has been included in intoDBP project, specifically in Work Package 1 as personnel costs.

4.2 Responsibilities for data management

The Project Coordinator has the ultimate responsibility for the data management in the project and will supervise the open science procedures during and after the project duration.

WP1 leader, will ensure dataset integrity and compatibility for its use during the project lifetime by different partners.

Each partner that generates data in the WP is responsible for registration and validation of datasets and metadata and must respect the policies set out in this DMP. Datasets

must be created, managed, and stored appropriately and in line with applicable legislation.

The detailed functions and responsibilities of the project coordinator, partners, and WP leaders will be described in the updated versions of the DMP in Month 36 (D1.2) and Month 48 (D1.3) of the project, respectively.

5.0 DATA SECURITY

5.1 Data recovery, secure storage, and transfer of sensitive data

Data collected from the partners will be digitised and stored on their own storage services. Only employees working for the project will have access to the data, using their username and password to access the files. The partners will also keep a back-up copies storage. In addition, the publicly available deliverables will be hosted on the server's website, following the host's security protocols.

The security and backup policy of the storage facilities of each partner will be described in the updated versions of the DMP in Month 36 (D1.2) and Month 48 (D1.3) of the project, respectively.

Table 3. List of data storage and recovery of all partners.

Partner	Data recovery and secure storage
Fundació Institut Català de Recerca de l'Aigua (ICRA)	Data collected from ICRA research groups will be digitized and stored on the file local server and backup copies are done to ensure the recovery of the data.
KIOS Center of Excellence– University of Cyprus (UCY)	The UCY is using the cloud-based Microsoft Office 365 suite. Due to the online and widely accessible nature of OneDrive, this storage could be used for open access data, restricted data with possible access from external collaborators, and internal data that can reside on remote Microsoft servers. A site/group collection in Office 365 SharePoint, groups together members and allows them to share, collaborate, and communicate over documents and tasks. In addition, sensor data received from external agencies will be stored in KIOS internal database file server. To ensure the recovery of the data, backups should be created in a regular basis. In KIOS CoE backups on virtual tape libraries physically stored on hard disc drives and actual tape libraries are maintained.
NIREAS International Water Research Centre – University of Cyprus (UCY)	UCY, and thus Nireas-IWRC and Chemistry Department, is using the cloud-based Microsoft Office 365 suite, which offers various online collaboration and

	storage tools (i.e., Outlook – email, Word/Excel/PowerPoint -- office applications, OneDrive and Teams). Data stored in Microsoft Office 365 comply with all the security and protection rules of the EU, as proper agreements were signed between the company and the European commission. Due to the online and widely accessible nature of OneDrive and Teams, these storage tools can be used for open access data, restricted data with possible access from external collaborators, and internal data that can reside on remote Microsoft servers. A site/group collection in Office 365 Teams, groups together members and allows them to share, collaborate, and communicate over documents. To ensure the recovery of the raw data even after complete destruction of the main data repositories, backups are being created in a regular basis via hard disc drives.
Dundalk Institute of Technology (DkIT)	Data collected from DkIT research group will be digitized and stored on file on the local server which is regularly backed up and on DkIT's OneDrive for Business cloud online file storage. Data will be archived on the Marine Institute and DkIT Open Access Repositories. Data stored long-term in the Marine Institute and DkIT Open Access Institutional repositories are protected against unauthorized access. Files on DkITs network and OneDrive will not be shared with any unauthorized people. DkIT have approved OneDrive for Business for the storage of sensitive and confidential information. Approval will be required to share any data with external colleagues.
Fundación Privada Instituto de Salud Global Barcelona (ISGLOBAL)	Data collected from the partners will be digitised and stored in secure servers that undergo regular backups, and protected with a password so only authorised study personnel will have access to preserve data protection. Data confidentiality and integrity are implemented at various levels: data storage and appropriate access levels, data in transit, and data access.
Aquasoil SRL (AS)	Data collected from AquaSoil will be digitized and stored on Google Drive folder, on local server and also backup copies (in hard disk) storage will be done. Data will be protected against unauthorized access with dedicated access levels.
Scan GmbH (SC)	All data is stored on s::can servers with regular automatic backups. Confidential data will be stored on

	separate network shares that are only accessible by people working on the project. Data shared within the consortium will be made accessible through an online server with appropriate user access levels.
Universita degli Studi di Catania (UNICT)	The data obtained from UNICT research groups will be converted into a digital format and stored on its respective storage service. Only project employees will be granted access to the data, using their individual usernames and passwords to retrieve the files. Additionally, UNICT will maintain regular backup copies of the data for storage purposes.
Water Europe (WE)	WE uses the cloud-based Microsoft Office 365 suite for its operational needs. Within Office 365 SharePoint, a site/group collection is established to bring together members, enabling them to effectively share, collaborate, and communicate on documents and tasks.
The Hebrew University of Jerusalem (HUJI)	Data collected from both HUJI research groups will be stored on local cloud servers with appropriate backup copies. Access will be granted to intoDBP partners upon request.
SK Euromarket LTD (SK Euromarket)	SK Euromarket possesses two on-site servers: one for storing general documents and another specifically dedicated to drawings. Additionally, the company ensures data protection by maintaining external backup copies for each local server and utilizes cloud-based backup solutions. These backup systems are fortified with antivirus security programs to safeguard the data.
Canal de Isabel II Gestión SA (CANAL)	Data collected from Canal de Isabel II will be digitised and stored on the file local server and backup copies are done to ensure the recovery of the data.
Water Board of Lemesos (WBL)	Data collected for the intoDBP sensors will be collected and stored at the server where s::can shall store the data. Data gathered from their telemetry systems stored at our local server which is regularly backup and protected against unauthorized access and use.
Ente de Abastecimiento de Agua Ter-Llobregat (ATL)	Data gathered from ATL will be converted into digital format and stored on a server located within the premises. Additionally, backup copies will be created regularly to guarantee the data can be restored if necessary. Appropriate access levels will be granted to intoDBP partners upon request.

6.0 ETHICAL ASPECTS

6.1 Introduction

All the activities carried out under the intoDBP project comply with ethical principles and relevant national, EU and international legislation.

The tasks for intoDBP concern research activities, data modelling and personal data. The surveys include healthy volunteers in the case studies areas of Spain, Cyprus, and Ireland. Ethics clearance by the competent local or national ethical committees will be received prior to proceeding with this work package.

6.2 Human Ethics

The transfer of data on human subjects to the CORA.RDR repository is only considered when: informed consents, ethics approval and -when applicable - approval by local data protection authorities cover the purpose that the data are envisaged to be used intoDBP and allows transfer individual aggregated data to the CORA repository. All data that are transferred to the intoDBP repositories will be pseudoanonymized. The Data Owner is responsible for the pseudoanonymization process for ensuring that identifiable variables are not transferred to the CORA repository. Directly identifiable variables include - but are not limited to - national ID number, name, phone number, ZIP-code, e-mail address, address, geographical coordinates (at a resolution that risks identification).One shall also be aware that a combination of just of few indirect identifying variables (such as birth data, gender, and zip-code) can be used to identify a large portion of individuals on any dataset. In this context, the Data Owner/Data Provider shall only provide such variables at the lowest possible resolution that is necessary for analysis, e.g. district instead of zip-code; year of birth or age instead of birth date.

6.3 Personal Data Processing

The surveys will take place in CS areas of Spain, Cyrprus and Ireland. The recruitment of participants and personal data collection will be different in each CS. Datasets will be centralised and analysed statistically by ISGlobal.

6.3.1 Personal Data Processing in CS area of Spain

An external service provider will be responsible of the recruitment of participants and data collection.

RECRUITMENT OF PARTICIPANTS

A survey to 1000 adult citizens will be conducted through a global platform for digital survey-based research. Participants will be selected from the metropolitan area of Barcelona based on age, gender, and residence area in order to have a stratified sample representative of the adult population in Barcelona Metropolitan Area. Potential participants will be contacted via email containing a brief explanation of what the study is about and its importance. Those who agree to participate will be contacted via email containing a brief explanation of what the study is about and its importance. Those who

agree to participate will give their informed consent before answering the questionnaire by clicking the corresponding box in the information sheet. The volunteers will be able to download such document. This project will follow ethic guidelines of Revised Declaration of Helsinki in its last version of 2023.

DATA COLLECTION

An online questionnaire has been designed to gather information on water consumption habits, including the type of water consumed at home and outside, the use of filters, water usage for cooking, bathing, and swimming, as well as sociodemographic characteristics and risk perception factors. The collected data will be anonymized by the external service provider.

PERSONAL DATA MANAGEMENT

The anonymization of personal data will be implemented by the service external provider.

PROTECTION OF PERSONAL DATA

The personal data will be completely anonymized by the global platform provider for digital survey-based research. This company places the highest priority on managing Personal Data in compliance with EU General Data Protection Regulation (GDPR). GDPR Compliance Statement and stakeholder roles are presented in Annex II. Briefly, Cint, the global platform provider for digital survey-based research, has developed and implemented a compliance program that addresses GDPR requirements. Key components of the program include:

- Senior management awareness, sponsorship, and regular compliance team reporting
- PIA and DPIA process
- Data Protection Officer appointment
- Regular privacy notice review and update
- Consent engine supporting consent by purpose
- Subject Access Request (SAR) process
- GDPR-compliant contract frameworks
- Breach notification
- Data Protection by Design
- Information Security controls
- Staff training
- On-going program governance

6.3.1 Personal Data Processing in CS area of Ireland

The personal data processing in Ireland has not been developed yet. This will be provided by DkIT during the project implementation and will be documented in the 2nd and 3rd versions of this deliverable.

6.3.1 Personal Data Processing in CS area of Cyprus

The personal data processing in Cyprus has not been developed yet. This will be provided by UCY during the project implementation and will be documented in the 2nd and 3rd versions of this deliverable.

7.0 OTHER

Some partners have their own procedures for data management or follow other procedures. These procedures are:

- KIOS Research and Innovation Center of Excellence of the University of Cyprus has its own Data Management Standard Operating Procedures.
- ISGlobal adheres to their own institutional procedure for data management.

8.0 REFERENCES

- DMP Online Tool from Catalan Open Research Agency: https://dmp.csuc.cat/
- Data Management in the context of Horizon 2020: http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

9.0 APPENDICES

9.1 Ethic issues

	yes, no or n/a
Humans	
I confirm that training certificates/personal licenses of the staff involved in animal experiments have been obtained and will be kept on file.	n/a
I confirm that templates of the informed consent forms and information sheets (in language and terms intelligible to the participants) will be kept on file.	Yes
Animal	
I confirm that relevant authorisations for animal experiments (covering also the work with genetically modified animals, if applicable) have been obtained, and will be kept on file.	n/a
Environmental protection and safety	
I confirm that appropriate health and safety procedures conforming to relevant local/national guidelines/legislation are followed for staff involved in this project.	Yes
I confirm that authorisations for relevant facilities (e.g. security classification of laboratory, GMO authorisation) have been obtained, and will be kept on file.	Yes
Third country	
I confirm that the research performed outside the EU is compatible with the Union, National and International legislation and could have been legally conducted in one of the EU Member States.	Yes
Data protection	
I confirm that a Data Protection Officer (DPO) has been appointed and the contact details of the DPO are made available to all data subjects involved in the research.	n/a
I confirm that data intended to be processed is relevant and limited to the purposes of the research project (in accordance with the 'data minimisation' principle).	Yes
I confirm that relevant authorisations for further processing of previously collected personal data have been obtained and will be kept on file.	Yes
I confirm that the data used are publicly available.	n/a

9.2 List of datasets

The list of datasets will be completed in the updated versions of the DMP in Month 36 (D1.2) and Month 48 (D1.3) of the project, respectively.

Type of Data (according to Table 1)	Description	Utility	Origin	Public (P) or Sensitive (S)	Generated (G) or Existing (E) data	Format	Size

intoDBP will create innovative tools and strategies to improve water quality management for safe human use and a healthy environment. It focuses on catchment protection and forecasting, transformative drinking water treatment, and real-time monitoring to combat the effects of climate and global change. In particular, intoDBP focuses on pollution and risks related to disinfection byproducts (DBPs). By developing and applying advanced, integrated, and cost-effective sensors and analytical methods, intoDBP will expand knowledge on water quality and DBP precursors to better understand its formation and human exposure in Europe. intoDBP monitoring results will feed into numerical forecasting tools to predict source water changes and formulate climate change adaptation pathways at catchment and treatment scale. intoDBP also develops transformative options for advanced and cost-effective upgrade of water treatment and disinfection.

In the intoDBP consortium researchers, small and large enterprises, communication experts and public services join forces to generate interdisciplinary solutions, that will generate a renewed perspective of drinking water surveillance, support decision-making and governance, and increase system resilience. intoDBP will implement and validate its cross-cutting products in four complementary case studies from three European countries where compliance with DBP regulation currently is an acknowledged challenge. The direct and visible positive impact of intoDBP in the case studies will foster rapid product adoption at a European and global scale, thus strengthening Europe's position and role in the global water market.

Reaching out beyond the water sector itself, intoDBP will directly engage society through surveys to analyze exposure to DBPs, collect data about catchment protection initiatives, create awareness and promote sustainable consumer behavior such as reducing bottled water consumption.

Funded by
the European Union