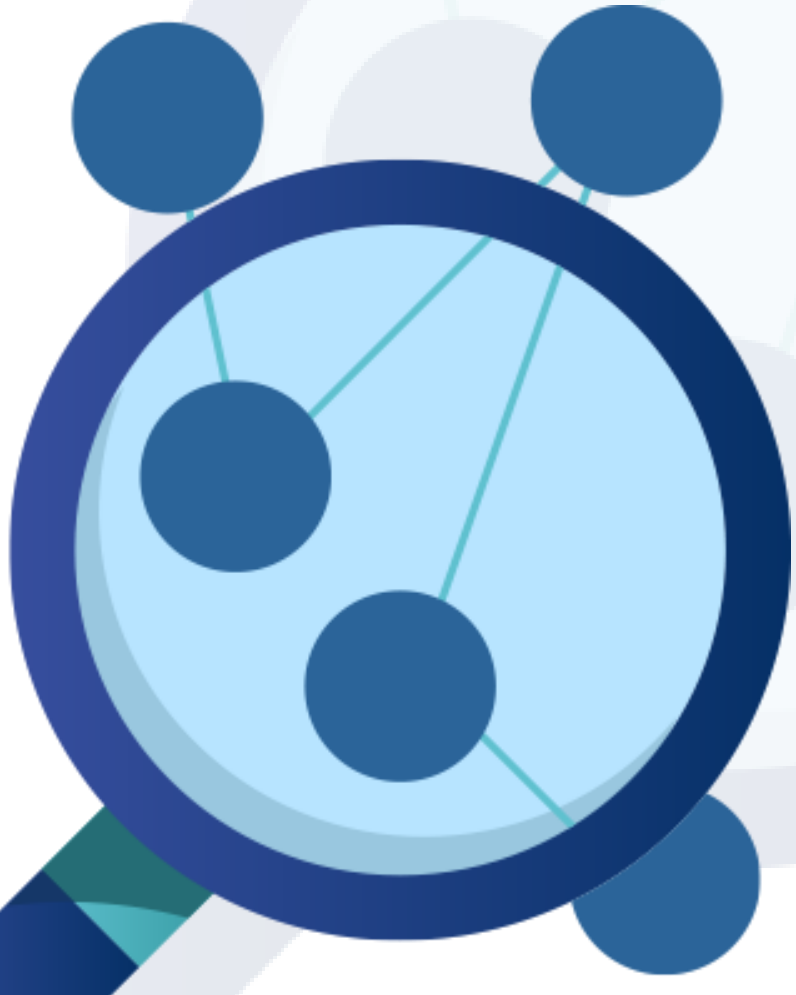
# Title: R1 Data management plan

Deliverable 1.2

WP 1 MANAGE- coordination and project management



Author: Sofia Semitsoglou Tsiapou, ICRA

Date: 28/11/2025

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 1.2** | | |
| **WORK PACKAGE** | **WP1** | | |
| **CONTRACTUAL DATE OF DELIVERY** | **30.11.2025** | | |
| **ACTUAL DATE OF DELIVERY** | **28.11.2025** | | |
| **NATURE** | **OTHER** | **DISSEMINATION LEVEL** | PUBLIC |
| **LEAD BENEFICIARY** | **ICRA** | | |
| **RESPONSIBLE AUTHOR** | SOFIA SEMITSOGLOU TSIAPOU (ICRA) | | |
| **CONTRIBUTIONS FROM** | | | |
| **ABSTRACT** | This Data Management Plan (DMP) outlines the principles, procedures, and tools applied to ensure effective management of all data generated, collected, or reused within the intoDBP project. It describes the types of datasets produced across the consortium, their origin, expected formats, size, and intended use, as well as the methods for ensuring that all data follow the FAIR principles—Findable, Accessible, Interoperable, and Reusable. The DMP defines data handling workflows, storage and backup procedures, access levels, ethical and legal considerations, and the responsibilities assigned to project partners. As a living document, this DMP will be updated throughout the project's lifetime to reflect the evolution of datasets, improvements in data workflows, and changes in project needs. This version (D1.2) corresponds to the update delivered at Month 36 of 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This deliverable contains original unpublished work except where clearly indicated otherwise.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 ABBREVIATIONS ..... 6

LIST OF FIGURES ..... 6

LIST OF TABLES ..... 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 current DMP ..... 8
1.4 Structure of this document ..... 9

2.0 DATA SUMMARY ..... 10

2.1 Purpose of the data collection/generation and relation to the project objectives 10
2.2 Origin of the data ..... 10
2.3 Reused data ..... 11
2.4 Expected size of the data ..... 12
2.5 Data utility ..... 12

3.0 FAIR DATA ..... 13

3.1 Making Data Findable including provisions for metadata ..... 13
3.1.1 Naming conventions ..... 13
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 ..... 17

4.1 Costs for making your data FAIR and long-time preservation ..... 17
4.2 Responsibilities for data management ..... 17

5.0 DATA SECURITY ..... 18

5.1 Data recovery, secure storage and transfer of sensitive data ..... 18

6.0 ETHICAL ASPECTS ..... 21

6.1 Introduction ..... 21
6.2 Human Ethics ..... 21
6.3 Personal Data Processing in CS area of Spain, Ireland and Cyprus ..... 21

7.0 OTHER ..... 23

8.0 REFERENCES ..... 24

9.0 APPENDICES ..... 25

9.1 Ethic issues ..... 25
9.2 List of datasets ..... 26

## LIST OF ABBREVIATIONS

| ABBREVIATION | DESCRIPTION |
|---|---|
| **CS** | Case Study |
| **DBP** | Disinfection By-Product |
| **DMP** | Data Management Plan |
| **DOM** | Dissolved Organic Matter |
| **EEM** | Excitation-Emission Matrix |
| **FAIR** | Findable, Accessible, Interoperable and Reusable |
| **GDPR** | General Data Protection Regulation |
| **GIS** | Geographic Information System |
| **HAZOP** | Hazard and Operability Assessment |
| **HRMS** | High-Resolution Mass Spectrometry |
| **ISIMIP** | Inter-Sectoral Impact Model Intercomparison Project |
| **SAR** | Subject Access Request |
| **THM** | Trihalomethanes |
| **UV** | Ultraviolet |
| **WP** | Work Package |

## LIST OF FIGURES

Figure 1. Data life cycle ..... 8

## LIST OF TABLES

Table 1. List of every type of data and its description. ..... 10
Table 2. List of Open Data repositories ..... 14
Table 3. List of data storage and recovery of all partners. ..... 18
Table 4. Ethic issues involved in the intoDBP project. ..... 25

## EXECUTIVE SUMMARY

This Data Management Plan (DMP) outlines how data generated within the intoDBP project will be collected, processed, stored, shared, and preserved in accordance with the FAIR principles and Horizon Europe requirements. It describes the types and origins of data produced across the consortium, as well as the procedures for ensuring data quality, metadata creation, interoperability, and secure storage. The DMP also specifies the responsibilities of partners in managing sensitive and personal information in compliance with ethical and legal standards.

This document is the updated DMP (D1.2), delivered at Month 36, and reflects current data management practices. It will be further refined in the final version (D1.3) at Month 48.

# 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 intoDBP 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.

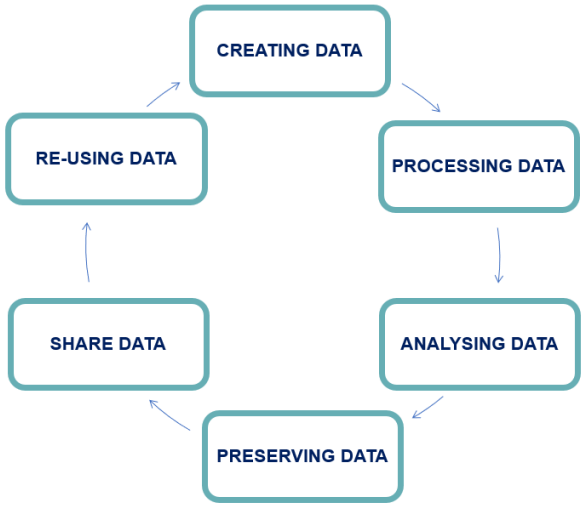



*Figure 1. Data life cycle*

## *1.2 Purpose of this document*

The purpose of deliverable D1.2: Data Management Plan is to describe the continuous progress of the data management plan (DMP), which was defined in D1.1, regarding the data generated by current intoDBP members for the duration of the project and all key elements for data handling and management during and after the project realization.

## *1.3 Objective of the current DMP*

The basic objective of a DMP document is to specify how data will be collected, processed, catalogued, and disseminated.

The DMP is not a fixed document but will be evolved as the project develops. This document represents an updated version (R1) 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.
- Section 9 lists the appendices, where the ethical aspects applying to the intoDBP project are detailed and the tables with the dataset information from every partner involved are demonstrated.

## 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*

In intoDBP project, the following type of data are currently being 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 data is provided by the Work Package Leaders during the project implementation and is documented in the current (i.e., 2nd) and 3rd version of this deliverable.

**Formats of the data**

The formats of the files so far generated and collected are:

- Graphics: jpg, jpeg, png, tif, tiff, gif,
- Tables/Data: csv, xlsx, excel, dat, hdf, shp, gpkg, txt, m, mat
- Text: docx, pdf, txt, ppt, pptx, docx and pdf
- Software: py, r scripts, m, mat, msx, inp
- Audio-visual: mp4
- Others: various raw formats

### *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 were either already re-used or will be re-used in the upcoming months:

| Data to be re-used | Respective deliverable and status |
|---|---|
| Historical data of UV sensors measurements of dissolved organic matter (DOM) to develop an early warning detection tool of DBPs (Task 3.4). | D3.4 – to be submitted on M46 |
| Historical meteorological data and DOM data together with archived weather forecast data to | D5.1 – submitted on M36 |

| | |
|---|---|
| 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). | D5.2 – to be submitted on M46 |
| Historical data on water use collected by on-going epidemiological studies in Europe (Task 6.1). | D6.1 – submitted on M36 |
| Data generated at WP 2 to develop a model for each DBP to advance the knowledge on exposure to DBP mixtures (Task 6.2). | D6.3 – to be submitted on M48 |
| Long-time series of monitored water quality parameters and climate variables to describe temporal patterns of THM concentrations and the climate variables (Task 6.3). | D6.2 – to be submitted on M46 |

### *2.4 Expected size of the data*

The data size varies significantly, based on the type of data generated by the partners. Data size of any type of data generated so far has been reported to be as small as 3MB (Office files) and as big as 5TB (Raw R&D data files). From those files, a selection will be uploaded to the respective repositories, where raw data, simulation data, scripts/algorithms will be omitted, leading to a few GB as an average of the deposited data.

### *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, demonstrating the impact potentials of the technology.

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 cases where private data are processed, permission will be requested by the provider prior to making the data publicly available.

Two different access levels will be defined:

- Open access refers to level of access with no restriction. It means that the information, data or resources are available to anyone without any restrictions or limitations. There are no special requirements necessary to access the information.
- Restrictive access applies when data cannot be openly shared due to legal, ethical, commercial, or security considerations. Datasets are available only under specific conditions, such as signing a data-sharing agreement, meeting eligibility criteria, or obtaining explicit authorization. Within this category, confidential access represents a stricter form of restriction, where data contain sensitive, private, or classified information, and access is granted only to authorized individuals. Under EU data-management principles (“as open as possible, as restricted as necessary”), restrictive and confidential access mechanisms ensure responsible management of sensitive datasets while still enabling controlled scientific reuse when appropriate.

Once processing, quality control, organization, 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 programs, 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 has received personal data in questionnaires, and such data were anonymized before making them available outside the project. A procedure for handling personal data is described 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

Since version D1.1, two partners have left the Consortium, Dundalk Institute of Technology (DkIT) and The Hebrew University of Jerusalem (HUJI). Researcher Valerie McCarthy moved from DkIT to Dublin City University (DCU) (Ireland), therefore DCU has now joined the Consortium. Two more partners, Helmholtz Centre for Environmental Research (UFZ) and Agencia Estatal Consejo Superior de Investigaciones Científicas (CSIC) have also joined. Therefore, the open data repositories, which are currently being used during the project, have been updated in Table 2 below:

*Table 2. List of Open Data repositories*

| **Name** | **Catalan Open Data 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. |

| | |
|---|---|
| Partners | University of Cyprus (UCY), Fundación Privada Instituto de Salud Global Barcelona (ISGLOBAL), Helmholtz Centre for Environmental Research (UFZ) and Agencia Estatal Consejo Superior de Investigaciones Científicas (CSIC) |
| **Name** | **OneDrive** |
| Website | https://onedrive.live.com/ |
| Description | OneDrive is a cloud storage service from Microsoft that allows users to store, sync, back up, and share files and photos across multiple devices. |
| Partners | University of Catania (UNICT) |
| **Name** | **NORMAN bioactivity database** |
| Website | https://www.norman-network.com/nds/ |
| Description | In the NORMAN Bioactivity database, activity/toxicity data for individual water relevant chemicals is collected from bioassays that measure endpoints relevant for water quality monitoring. It is part of the NORMAN network, which enhances the exchange of information on emerging environmental substances, and encourages the validation and harmonisation of common measurement methods and monitoring tools. |
| Partners | Helmholtz Centre for Environmental Research (UFZ) |
| **Name** | **Google Drive** |
| Website | https://drive.google.com/ |
| Description | Google Drive is a cloud-based file-hosting service developed by Google that allows users to store, synchronize, and share files online. |
| Partners | Aquasoil SRL |
| **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 S:can Iberia |
| **Name** | **DCU Open Data Repository (DORAS)** |
| Website | https://doras.dcu.ie/ |
| Description | DORAS is DCU's new Open Access Institutional Repository, providing a secure central repository for the research output of the university. DORAS makes scholarly information produced in DCU available to the wider academic community and offers greater visibility and higher impact for this material which can be globally accessed over the internet. The development of DORAS is underpinned by the principles of the Open Archive Initiative. |
| Partners | Dublin City University (DCU) |

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), SK Euromarket and Eparchiakos Organismos Aftodioikisis Lemesou (WBL) will be updated in the next version of this DMP (D1.3). 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.

## 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 updated information regarding security and backup policy of the storage facilities of each partner is described in Table 3, with the final, updated version to be prepared in M48 (D1.3).

*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<br><br>NIREAS International Water Research Centre<br><br>University of Cyprus (UCY) | The UCY, and thus KIOS CoE, 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, and OneDrive). 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, 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 even after complete destruction of the main data repositories, backups should be created in a regular basis. In KIOS CoE we maintain backups on virtual tape libraries physically stored on hard disc drives and actual tape libraries. |

| | |
|---|---|
| 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 in order to preserve data protection. |
| Aquasoil SRL (AS) | Data collected from AquaSoil research staff will be digitized and stored on DRIVE folder, on local server and also backup copies (in hard disk) storage will be done. AS will maintain regular backup copies of the data. |
| Scan GmbH (S::CAN) | 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) | UNICT is using the cloud-based Microsoft Office 365 suite, which offers several tools for storage and communication. Data are stored in Microsoft Office 365 using OneDrive application, complying 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, 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. |
| 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. |
| 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 will be digitized and stored on the file local server and backup copies are maintained. |
| Water Board of Lemesos (WBL) | Data collected for the intoDBP sensors will be collected and stored at the server where S::CAN stores 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. |
| Helmholtz Centre for Environmental Research (UFZ) | The user and group directories are backed up daily to tape (once a month as a full backup, weekly as a cumulative backup and daily as a differential backup). The retention period is 6 weeks for a full backup, 2 weeks for a cumulative backup and 10 days for a differential backup. Data archiving is done via UFZ internal Data management portal or Omero (Open Microscopy Environment Remote Objects), which is an open source client/server system written in Java for visualizing, managing, and annotating microscope images and metadata. |
| Dublin City University (DCU) | Data collected within the DCU research group will be stored on secure institutional servers with controlled access. A mirrored back-up is maintained through DCU's internal storage systems, and working copies are stored in a restricted-access Google Drive folder to ensure redundancy. Sensitive or newly collected data are only accessible to authorised project members, following DCU's Data Governance and Data Protection policies. |
| Agencia Estatal Consejo Superior de Investigaciones Científicas (CSIC) | Local backup copies are maintained. |

## 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 human studies 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*

All data that are transferred into the intoDBP repositories are anonymized. The data creator is responsible for the anonymization process, which is performed by the global platform provider Cint (www.cint.com), 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 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 Personal Data Processing in CS area of Spain, Ireland and Cyprus*

Population surveys took place in CS areas of Spain, Cyprus and Ireland. The recruitment of participants and personal data collection was different in each CS. Cint was the external service provider responsible for the recruitment of participants and data collection. Datasets were centralised and analysed statistically by ISGlobal.

A cross-sectional study was conducted in the Barcelona Metropolitan Area (Spain), Cyprus and the Republic of Ireland. Sample size was calculated separately in the three study sites using a survey specific formula. The number of participants was 1000 in Barcelona, 412 in Cyprus, and 532 in Ireland. Recruitment of study participants was conducted through the subcontracted international survey platform (www.cint.com), who contacted volunteers in their databases. Sampling was stratified, to enroll participants representative of the population in the study base

by age, gender and geographical area distribution. An online questionnaire was 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 study was approved by the Ethics Committees of the respective study areas or countries and participants provided informed consent to participate before enrollment.

The results of this study have been finalized in the form of deliverable D6.1 (completed on M36).

## 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*

*Table 4. Ethic issues involved in the intoDBP project.*

| | 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

A current list of datasets for the period M1-M36 is given below. The datasets of Water Europe (WE), SK Euromarket (SKE) and Eparchiakos Organismos Aftodioikisis Lemesou (WBL) and any existing information will be updated in the final version of this DMP (D1.3).

| ICRA | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type of Data (according to Table 1) | Description | Utility | Origin | Public (P) or Sensitive (S) | Generated (G) or Existing (E) data | Format | Size |
| Project management | Meeting Minutes and Data Management Plan | Project Management | Meetings | P | G | .docx | 5 GB |
| Experimental | Results (Concentrations) | Data to be processed | Method development | P | G | .xlsx | 5 GB |
| Metadata | Metadata | Share the Data within the consortium | Complementary data | P | G | xlsx | 10 GB |
| Methodology | Scripts and algorithms | Fingerprinting - Used for data processing and for modelling workflows | Used for data processing and for modelling workflows | P | G | R scripts | 10 GB |
| Dissemination | Papers | Dissemination of intoDBP activities | Research data | P | G | .docx and pdf | 10 GB |

| SCAN | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type of Data (according to Table 1) | Description | Utility | Origin | Public (P) or Sensitive (S) | Generated (G) or Existing (E) data | Format | Size |
| Experimental | spectra, time series | Analyze for DBP concentration | S:CAN measurement equipment | S | G | Text | N/A |
| Experimental | spectra, time series | Analyze for DBP concentration | S:CAN measurement equipment | S | G | Text | N/A |

| AS | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type of Data (according to Table 1) | Description | Utility | Origin | Public (P) or Sensitive (S) | Generated (G) or Existing (E) data | Format | Size |
| Design | files | MITO3X development: design | Design data | P | G | pdf & dwg | 20 GB |
| Project Magement | files | MITO3X development: part quote | Design data | P | G | pdf & dwg | 10 GB |
| Experimental | Office files | Construction and operation of pilots | R&D data | P | G | Raw files | 10 GB |
| Experimental | Office files | Results | R&D data | P | G | Raw files | 5 TB |
| Dissemination | Office files | Demonstration | R&D data | P | G | Excel/ jpeg | 10 GB |
| Experimental | Office files | Implementation of fluoro-absorbance sensors for process optimization | R&D data | P | G | Excel | 10 GB |
| Dissemination | Office files | Demonstration of transformative approaches for DBPs minimization | R&D data | P | G | Excel | 1 GB |
| Dissemination | Papers | Dissemination of intoDBP activities | Research data | P | G | docx and pdf | 10 GB |

| ATL | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type of Data (according to Table 1) | Description | Utility | Origin | Public (P) or Sensitive (S) | Generated (G) or Existing (E) data | Format | Size |
| Observational | observational data | data from S:can sensors installed in Ter DWTP | sensors | S | G | xlsx | 5-10 MB |
| Experimental | experimental data | water quality data to manage DWTP | ATL laboratories | S | G | xlsx | 5-10 MB |
| Observational | historical data | operational historical data | ATL databases | S | E | xlsx | 5-10 MB |
| Observational | historical data | DBPs exposure data | ATL databases | S | E | xlsx | 5-10 MB |
| Observational | observational data | data from S:can sensors installed in Ter DWTP | sensors | S | G | xlsx | 5-10 MB |

| CANAL | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type of Data (according to Table 1) | Description | Utility | Origin | Public (P) or Sensitive (S) | Generated (G) or Existing (E) data | Format | Size |
| Observational | Water quality | Canal de Isabel II | Own sources | S | G | .xlsx | 3MB |
| Observational | Distribution network data | Canal de Isabel II | Own sources | S | E | .xlsx | 3MB |
| Observational | Water quality | Canal de Isabel II | Own sources | S | G | .xlsx | 3MB |

| CSIC | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type of Data (according to Table 1) | Description | Utility | Origin | Public (P) or Sensitive (S) | Generated (G) or Existing (E) data | Format | Size |
| Simulation | Simulation | Production of scientific publication | Obtained from models, codes used to produce the data are fully open | P | G | R, Python, CSV, PDF | 150MB |

| DCU | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type of Data (according to Table 1) | Description | Utility | Origin | Public (P) or Sensitive (S) | Generated (G) or Existing (E) data | Format | Size |
| Observational | High Frequency Data | Raw Sensor High Frequency Monitoring Data (Marine Institute) | Marine Institute | P | E | .csv | > 8 GB |
| Derived or compiled | High Frequency Data | Processed Sensor High Frequency Monitoring Data (Marine Institute) | Marine Institute | P | E | .csv | > 8 GB |
| Observational | High Frequency Data | Met Eireann Meterological Data | Met Eireann | P | E | .csv | > 1 GB |
| Observational | Water Quality Data | In Situ Water Quality Data | Newly Collected | P | G | .xls, .csv | > 1 GB |
| Observational | Sensor Data | Newly Collected Sensor Data | Newly Collected | S | G | .xls, .csv | > 1 GB |
| Observational | Survey Data | Survey Data on Water Use and Behaviour | Newly Collected | P | G | .xls | <1 GB |
| Simulation | High Frequency Data (subdaily) | European Centre for Medium-Range Weather (ECMWF) Forecasts | ECMWF | P | E | .xls, .csv | < 8 GB |
| Simulation | High Frequency Data (subdaily) | NOAA daily forecasts | NOAA | P | E | .xls, .csv | < 8 GB |

| ISGlobal | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type of Data (according to Table 1) | Description | Utility | Origin | Public (P) or Sensitive (S) | Generated (G) or Existing (E) data | Format | Size |
| Observational | Observational data | Working baseline to produce the scientific publications | Responses of online questionnaire about water use and consumption in the CS sites; data on water use collected by on-going epidemiological studies in Europe. | S | G/E | xlsx, R | 1GB max |
| Experimental | Experimental data | Working baseline to produce the scientific publications | Data of concentrations of the targeted DBPs and other drinking water parameters generated in WP2, Task 2.1; long time period of monitored water quality parameters and climate variables provided by the water utilities in Barcelona (ATL), Limassol (WBL), and the Group Water Schemes, Environmental Protection Agency, and other sources in Ireland. | S | G/E | xlsx, other | 1GB max |
| Simulation | Simulation data | Working baseline to produce the scientific publications | Obtained from predictive models of emerging DBPs based on routinely monitored parameters in selected case-studies in task 6.2. | S | G | R, other | 1GB max |
| Derived or compiled | Derived or compiled data | Working baseline to produce the scientific publications | Obtained from data mining and statistical analysis of all the observational and experimental data generated or already existing for tasks 6.1., 6.2., 6.3. | S | G | R, other | 1GB max |
| Metadata | Metadata | Working baseline to | Information about all the generated or existing | S | G/E | xlsx, docx, | 1GB max |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | produce the scientific publications | data. | | | pdf, R… | |
| Methodology | Methodology data | Working baseline to produce the scientific publications | Data from standard operation procedures or technical specifications of instrument/process. | S | G | xlsx, docx, pdf… | 1GB max |
| Project management | Project management data | Working baseline to produce the scientific publications | Generated during the project's implementation by the intoDBP consortium. | S | G | xlsx, docx, pdf… | 1GB max |
| Dissemination | Dissemination data | Working baseline to produce the scientific publications | Obtained from all dissemination and communication activities of intoDBP project. | P | G | jpeg, png, pdf, pttx, xlsx, docx, pdf… | 1GB max |

| UCY-KIOS-NIREAS | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type of Data (according to Table 1) | Description | Utility | Origin | Public (P) or Sensitive (S) | Generated (G) or Existing (E) data | Format | Size |
| Observational | GIS Data | WBL/UCY | WBL | S | E | .shp, gpkg, PostgreSQL | 15 GB |
| Methodology | Source Code | UCY | UCY | P | G | .m, .py | 500 MB |
| Observational | Sensor Data | WBL/UCY | WBL/SCAN | S | G | PostgreSQL | 2 GB |
| Simulation | Models | intoDBP partners and case study boards | UCY/WBL/SCAN | P | G | .msx,.inp, .shp, .png, .txt, .xlsx, .pdf, .gif, .mp4 | 100 MB |
| Experimental | Laboratory Data | UCY/Nireas/Chemistry | UCY Laboratory | S | G | xlsx, .csv | 50-200MB |
| Other | Other Data | UCY | UCY/intoDBP partners | S | G | .pdf, .txt, .jpg, .png, .docx, .xlsx, .pptx | 500 MB |

| UFZ | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type of Data (according to Table 1) | Description | Utility | Origin | Public (P) or Sensitive (S) | Generated (G) or Existing (E) data | Format | Size |
| Experimental | Raw data from in vitro assays | In vitro bioassay data | Incucyte S3, Imagexpress microscope and Tecan Reader measurements | S | G | Various raw data formats from Incucyte S3, Imagexpress microscope and Tecan Reader | 1 TB |
| Derived or compiled | Processed data from in vitro assays | In vitro bioassay data | Raw data processed with R-workflow for assay evaluation (https://git.ufz.de/braung/automatedbioassayscreening) | S | G | Excel and PDF | 5 GB |
| Derived or compiled | Inhibitory concentrations or effect concentrations for assay endpoints | Sample effect concentrations | Raw data processed with R-workflow for assay evaluation | S | G | Excel | 1 GB |

| DCU | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type of Data (according to Table 1) | Description | Utility | Origin | Public (P) or Sensitive (S) | Generated (G) or Existing (E) data | Format | Size |
| Experimental | Fluorescence and absorbance spectra | Development of prototype fluorescence tool (D 2.3) - Comparison EEM/HRMS fingerprinting - Fluorescence based prediction tool for DBPs using sensors data from samples | Sampling campaigns from all case studies | S | G | .csv,.xslx,, .txt, .pdf, .ppt | 2.6 GB |
| Methodology | Matlab scripts | Development of prototype fluorescence tool (D 2.3) - Comparison EEM/HRMS fingerprinting - Fluorescence based prediction tool for DBPs using sensors data from samples | Sampling campaigns from all case studies | S | G | .csv,.xslx, .m, mat, .txt, .pdf, .ppt | > 8 GB |
| Observational | Spectral and physicochemical parameters | Online prediction of DBPs using real-time fluorescence data from sensors installed at monitoring stations of water distribution system of Limassol | Online data from web platform provided by SCAN | S | G & E | .csv,.xslx, .m,.mat, .txt, .pdf, .ppt | 100 MB |
| Dissemination | Text documents, excel sheets, pictures, PPT slides | Production of documentation: Experimental Plan of research activities at the pilot water treatment unit of Limassol (Cyprus); Document for HAZards and Operability (HAZOP) assessment related to such research activities | Literature review | S | G | .docx, .txt,.xslx, .pdf, .ppt, .JPEG | 160 MB |
| Experimental | Text documents, excel sheets, | Production of documentation: Experimental Plan of research activities at the pilot water treatment | Jar tests conducted by UCY | S | G | .docx, .txt,.xslx, .pdf, .ppt, | > 1 GB |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | pictures, PPT slides | unit of Limassol (Cyprus); Document for HAZards and Operability (HAZOP) assessment related to such research activities | | | | .JPEG | |
| Methodology | Text documents, excel sheets, pictures, PPT slides | Production of documentation: Experimental Plan of research activities at the pilot water treatment unit of Limassol (Cyprus); Document for HAZards and Operability (HAZOP) assessment related to such research activities | Safety documentation of chemicals and materials used | S | G | .docx, .txt,.xslx, .pdf, .ppt, .JPEG | <1 GB |

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.



intoDBP

Funded by
the European Union