



# Data Management Plan D7.2

Authors: Murray, A.M., Comeau, L.E.L., Lørup, J.K.



## Disclaimer

This report was written as part of the PHISHES project under EC grant agreement 101157438. The information, documentation and figures available in this deliverable were written by the PHISHES project consortium and do not necessarily reflect the views of the European Commission. The European Commission is not liable for any use that may be made of the information contained herein.

## Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise.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.

## How to quote this document?

Murray, A.M., Comeau, L.E.L., Lørup, J.K. (2025). Data Management Plan (PHISHES Deliverable 7.2)



This deliverable is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).



Project Acronym: **PHISHES**

Project Full Title: **Physically-Based Integrated Soil HEalth Simulation Platform**

Grant Agreement Number: **Project: 101157438 - HORIZON-MISS-2023-SOIL-01**

Type of Action: **HORIZON-RIA**

<b>Deliverable number</b>	<b>7.2</b>
Deliverable name:	Data Management Plan
Dissemination level:	Public
Type of document:	Report
Deliverable Leader:	DHI
Delivery due date:	28 <sup>th</sup> February 2025
Actual date of submission:	28 <sup>th</sup> February 2025
Deliverable description:	This deliverable provides the initial plan for management of the data generated and collected during the PHISHES project. The DMP describes the data management life cycle for all datasets to be collected, processed, and/or generated.
WP / WP number:	WP7
Lead author(s):	Alexandra Murray (DHI), Laura Comeau (DHI), Jens Kristian Lørup (DHI)
Contributors:	Alexandra Murray (DHI) Jens Kristian Lørup (DHI) Laura Comeau (DHI) Roar Askær Jensen (DHI) Robert Earon (SGI) Charlotta Tingberg (SGI) Václav Rára (CG) Devau Nicolas (BRGM) Katja Firus (T6)
Reviewers:	All partners



## Version log

Version	Date	Issued by	Description	Summary of changes
1	10.02.2025	Alexandra Murray (DHI)	Initial draft for internal review	
2	27.02.2025	Laura Comeau (DHI)	Second version for submission	Integration of suggestions from partners



## Table of Content

List of Tables .....	6
Executive Summary .....	7
1. Introduction .....	9
2. Scope of the DMP and EU Requirements .....	10
2.1 Scope of the DMP .....	10
2.2 The EU requirement on DMP .....	10
2.3 FAIR Principles .....	11
3. Data Summary .....	13
3.1 Datasets .....	13
3.2 Data Reuse .....	19
4. FAIR Principles .....	20
4.1 Data Findable .....	20
4.2 Data Accessible .....	21
4.3 Data Interoperable .....	22
4.4 Data Reusable .....	23
5. Data Security .....	24
5.1 Data preservation and processing .....	24
6. Allocation of Resources .....	25
7. Privacy and Ethical Issues .....	26
8. Scientific Publications .....	27
8.1 Green open access .....	27
8.2 Gold open access .....	27
9. Conclusion .....	28
10. References .....	29



## List of Tables

Table 1. PHISHES Digital Platform (Work Package 1), Lead partner: DHI .....	14
Table 2. PHISHES Impact of land use management on fate and transport of contaminant in soils (Work Package 2), Lead Partner: Bureau de Recherches Géologiques et Minières (BRGM) .....	15
Table 3. PHISHES Test cases (Work Package 4.1), Lead Partner: CzechGlobe – Global Change Research Institute (CG) .....	16
Table 4. PHISHES Test cases (Work Package 4.2), Lead Partner: Warsaw University of Life Sciences (SGGW).....	17
Table 5. PHISHES Test cases (Work Package 4.3), Lead Partner: Swedish Geotechnical Institute (SGI).....	18
Table 6. Metadata standard elements for describing a resource and its attributes .....	21



## Executive Summary

Deliverable 7.2 - Data Management Plan (DMP) is a document that aims at providing rigorous detailed information and guidelines about the way data is managed within the PHISHES project.

The DMP describes the data management procedures that will be implemented in PHISHES according to Horizon Europe Programme standards considering Open Science Policy, FAIR (Findability, Accessibility, Interoperability and Reuse) Principles, and relevant EU laws.

The PHISHES project data policy should promote Data Sharing following the principle 'as open as possible as closed as necessary'. There are some types of data that are subject to specific limitations from the data provider (as stated in the Consortium Agreement and DMP) such as restricting public access due to requirements on anonymity. Before external publication or dissemination of any kind, including scientific dissemination, approval from stakeholders / data owners must be secured.

This document details the different input datasets that are exploited within the project, in addition to providing an overview about the output data generated through the implementation of the project. More so, this deliverable provides information related to the application of FAIR principles, data security, and the allocation of resources related to data management.

The DMP is a living document that will continuously be updated throughout the life of the project.

## Keywords

Data Management, Data Storage, Data Life Cycle, Management, Logistics.

## List of Acronyms and Abbreviations

<b>API</b>	Application Programming Interface
<b>BRGM</b>	Bureau de Recherches Géologiques et Minières, France
<b>CA</b>	Consortium Agreement
<b>CC</b>	Creative Commons
<b>CG</b>	CzechGlobe – Global Change Research Institute
<b>DMP</b>	Data Management Plan
<b>DOI</b>	Digital Object Identifier
<b>EC</b>	European Commission
<b>EU</b>	European Union
<b>EUDPR</b>	European Data Protection Regulation for Union
<b>EOSC</b>	European Open Science Cloud
<b>FAIR</b>	Findability, Accessibility, Interoperability and Reuse



<b>GA</b>	Grant Agreement
<b>GDPR</b>	General Data Protection Regulation
<b>GLA</b>	General Assembly
<b>IDF</b>	International DOI Foundation
<b>IPR</b>	Intellectual Property Rights
<b>ISO</b>	International Organization for Standardization
<b>OGC</b>	Open Geospatial Consortium
<b>ORE</b>	Open Research Europe
<b>PCT</b>	Project Coordination Team
<b>PDP</b>	PHISHES Digital Platform
<b>PHISHES</b>	PHysically-Based Integrated Soil HEalth Simulation Platform
<b>SGGW</b>	Warsaw University of Life Sciences
<b>SGI</b>	Swedish Geotechnical Institute
<b>URL</b>	Uniform Resource Locator

Acronyms and abbreviations within the dataset tables in Section 3.1 Datasets are written below each data table for ease of reference.





## 1. Introduction

This deliverable is the first version of the Data Management Plan (DMP) for the PHISHES project, established within the task “Research data, knowledge and IPR management”, which is led by DHI. The objective of the task is to develop and provide data management policies and recommendations that are concerned with the acquisition, storage, security, retrieval, dissemination, archival, and disposal of data collected within the project.

Included within this deliverable is the oversight and handling of the project Open Data policy, as well as management of third parties’ open science and/or data within the Consortium. The DMP will facilitate the different steps to be taken to adhere to the Open Science policy which will maximise access to, and use of research data generated by the project. This DMP will provide information on the data that will be collected, newly produced, and/or elaborated during the project by all partners, and to frame the project strategy concerning Open Data (defining how project-based data and tools will be handled and made transparent and open access).

Within the PHISHES project there are 5 test sites, each located in a different geographical location (Zelivka river basin, Czech Republic; Koprzywianka River, Poland; Ornskoldsvik, Sweden; BRGM (Bureau de Recherches Géologiques et Minières) test facility, France; Rotterdam urban area, Netherlands). To ensure appropriate collection of relevant datasets, enable technical partners to undertake relevant analysis, and ensure external stakeholder participation across the different sites, it is vital for the PHISHES project to establish clear and robust data management guidelines that will facilitate transparency and compliance with the Findability, Accessibility, Interoperability and Reuse (FAIR) data principles.

The PHISHES project data policy should promote Data Sharing following the principle ‘as open as possible as closed as necessary’. There are some types of data that are subject to specific limitations from the data provider (as stated in the Consortium Agreement and DMP) such as restricting public access due to requirements on anonymity. Before external publication or dissemination of any kind, including scientific dissemination, approval from stakeholders / data owners must be secured.

The DMP adopted within the PHISHES project strictly adheres to the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, on the protection of natural persons about personal data processing and on the free movement of such data, repealing Directive 95/46/EC – General Data Protection Regulation (GDPR). To this end, the scope of the DMP is to outline what and how data will be collected, processed, and/or generated, what methodologies and standards will be applied, whether data will be shared/made open access, and how data will be stored during and after the end of the Project.

The deliverable objectives are:

- Providing an overview of the input/output data that will be reused/generated within the project.
- Providing a summary of PHISHES datasets.
- Identifying relevant dataset attributes such as the type and format of the datasets, in addition to the methodologies used for data collection, analysis, and reuse in parallel with other deliverables.
- Setting up clear guidelines that project partners ought to follow to ensure that the work is performed according to the appropriate guidelines and in compliance with the FAIR guidelines.
- Defining a data management governance structure to mainstream these guidelines and to ensure satisfactory completion of the PHISHES Grant Agreement (GA).

All PHISHES partners have acknowledged the principles set forth in this DMP.



## 2. Scope of the DMP and EU Requirements

### 2.1 Scope of the DMP

The release of the DMP is part of the project task “Research data, knowledge and IPR management”. The objectives of this task are to oversee and handle the project open data policy, including management of third parties’ open science and/or data within the consortium, and to frame the project strategy concerning open data. The DMP will reflect the status of the Consortium Agreement (CA) on data management and is consistent with the exploitation and Intellectual Property Rights (IPR) requirements.

The DMP is intended to be a living document where information is continuously added and revised as the project implementation progresses and knowledge about data emerges and consolidates. This first release covers the most relevant aspects of data management and provides the following:

- A high-level mapping of the project data input sources and input datasets even if temporary in this version of the deliverable, as more detailed datasets (input/output) will be thoroughly identified in work packages “Impact of land use management on fate and transport of contaminant in soils” (Work Package 2), “Response of soil ecosystem functions/services to soil contamination” (Work Package 3), and “PHISHES test cases” (Work Package 4), including relevant dataset attributes.
- Guidelines directed towards ensuring data within PHISHES is FAIR and follows Horizon Europe Programme and EU Open Science policy.
- An overview about data preservation and security of data within the project.
- A summary of the different ethical issues and the processing of personal data within the project. The data governance structure is aimed at ensuring that the above-mentioned guidelines are properly streamlined and implemented.

The DMP is a living document that will be constantly updated throughout the life of the project in conjunction with the different related project deliverables that feed directly or indirectly the DMP. All the information that is included here will be further articulated to ensure the accuracy, reliability, and usability of this guide.

### 2.2 The EU requirement on DMP

The principles of openness and transparency will underlie all research activities in the PHISHES project in line with the open science policy and Horizon Europe guidelines (European Commission, 2024), encouraging the use of the Open Research Europe (ORE) publishing platform and the open repository for research objects. The objectives of open science policy under Horizon Europe are as follows:

- Ensuring that beneficiaries retain the intellectual property rights they need to comply with their open access obligations.
- Requiring research data to be FAIR and open by default (with exceptions notably for commercial purposes).
- Promoting the adoption of open science practices, from sharing research outputs as early and widely as possible, to citizen science, and developing new indicators for evaluation research and rewarding researchers.
- Engaging and involving citizens, civil society organisations, and end-users in co-design and cocreation processes and promoting responsible research and innovation.
- Promoting the usage of the European Open Science Cloud (EOSC).



Open access practices in Horizon Europe Open Access Guidelines (2017) are defined as follows:

- Open access to research outputs such as publications, data, software, models, algorithms, and workflows.
- Early and open sharing of research, for example through preregistration, registered reports, preprints, and crowdsourcing of solutions to a specific problem.
- Use of open research infrastructures for knowledge and data sharing.
- Participation in open peer-reviews.
- Measures to ensure reproducibility of results.
- Open collaboration within science and with other knowledge actors, including involving citizens, civil society, and end-users - such as in citizen science.

Concretely, according to the Regulation (EU) 2021/695 of the European Parliament (2021) and of the Council of 28 April 2021 establishing Horizon Europe – the Framework Programme for Research and Innovation, all Horizon Europe program beneficiaries, including all the partners of the PHISHES project, shall commit to the following:

1. Open access to research data shall be the general rule under the terms and conditions specified in the grant agreement, ensuring the possibility of exceptions following the principle “as open as possible, as closed as necessary”.
2. Beneficiaries shall manage all research data generated in an action under the Programme in line with the FAIR principles and in accordance with the grant agreement.

The four FAIR principles are to be implemented when relevant to govern research data and project outputs.

## **2.3 FAIR Principles**

The principles of Findability, Accessibility, Interoperability, and Reusability are set to serve and guide data producers and authors as they navigate around any obstacles related to the findability, accessibility, interoperability, and reusability of scientific data. The FAIR principles were designed to maximise the added value gained by scientific publications and findings (Wilkinson et al., 2016).

The principles are not only important to data in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. More so, FAIR data can leverage multiple benefits such as:

- Gaining maximum potential from data assets.
- Increasing the visibility and citations of research.
- Improving the reproducibility and reliability of research.
- Staying aligned with international standards and approaches.
- Attracting new partnerships with researchers, business, policy, and broader communities.
- Enabling new research questions to be answered.
- Using new innovative research approaches and tools.
- Achieving maximum impact from research.

FAIR principles serve as guidelines to support best practices in data management by researchers and data stewards and should be tailored to the specific context in which they are applied.



According to Wilkinson et al. (2016), the PHISHES project data should be FAIR as follows:

***Findable:***

- Each dataset should be uniquely identifiable.
- Data should be discoverable using standard methods, e.g., search engines that operate on keyword searches.
- Data should be discoverable by both humans and machines, e.g., humans doing internet searches and machines crawling for relevant metadata.

***Accessible:***

- Data should be accessible to anyone who needs it.
- Data should be accessible for the long term.

***Interoperable:***

- Data should not use closed or proprietary formats.
- Data should be described using standardised vocabularies, and the vocabularies' definitions should be referenced in the metadata.

***Reusable:***

- Data should be licensed in such a way that reuse is permitted.
- Data should be citable.

All beneficiaries of the Horizon Europe programme must commit themselves to the FAIR principles when it comes to the management of research data. The project GA highlights the FAIR principles and articulates how the principles were designed to serve the community as a minimal scope approach which focuses on the specification of minimally required standard protocols, lightweight interfaces, and formats. In addition, the GA also underlines that promoting FAIR data is all about communication, reconciliation, adoption, and endorsement of community standards for data provenance, versioning, identity, citation, and dependency.



### 3. Data Summary

A first version of a PHISHES data summary is presented in this chapter. The summary is guided by the Horizon Europe programme Data Management Plan Template (HE) (V1.1 – 01.04.2022).

The template has been used to prepare a Data Survey Form, which has been sent to the responsible persons for each of the work packages requesting an overview of the data they currently anticipate they will be using and producing during the project. The tables presented in this chapter are therefore the result of the feedback from the different work packages as well as the meetings and discussions, which have taken place within the project team.

This is the first of four versions of the Data Management Plan. At this early stage of the project, it is not yet well-defined which model simulations and specific scenarios will be run and which corresponding outputs will be produced.

The focus of this first version of the DMP has therefore been on the input data to be used in the project, e.g., datasets that will be used as input data to the modelling exercises as well as data used to calibrate and validate these models. It has also been decided to primarily list data which the partners know will be used in the project, rather than include data which is uncertain whether it will be used and therefore may have to be deleted from the next version of the report.

In the later versions of the DMP the output data and information will be included when there is a clearer picture of the extent, type and format of these output data.

Guided by the Horizon Europe template the data gathered has focused on the following types of information:

- A description of the type of data.
- The type and origin of the data source (e.g., data from existing and/or ongoing projects, data to be produced within the project, or data from other sources, e.g., farmers, government agencies, etc.).
- The purpose of the data within the project, i.e. who will be using the data and for what purpose.
- List of work packages, which will either use and/or produce the data. Some work packages may carry out test cases and other work packages may use these data for model setup and model calibration and validation.
- The data type, e.g., spatial data, time series data, time varying 2D data, etc.
- The data format (CSV, shapefiles, etc.) and size (GB) of the data.
- To which extent the data will be available outside of the project. There are some types of data that are subject to specific limitations from the data provider such as restricting public access due to requirements on anonymity.
- To whom outside of the project may the data be useful.

Apart from providing input to the Data Management Plan, the information gathered through the Data Survey Forms will be combined and shared between all partners and work packages in the project. It will therefore serve as a dynamic list, which will help provide an overview of the data being used and produced by the project and its different work packages. This will facilitate data sharing and optimize its use within the project.

#### 3.1 Datasets

As mentioned above, this first version of the DMP focuses on the input datasets used in the different work packages in the project. These are presented in the following tables, with a separate table for each of the work packages.



Table 1. PHISHES Digital Platform (Work Package 1), Lead partner: DHI

	Data Description	Data Source	Origin of Data	Data Type	Data format	Metadata available	Keywords available	Can data be available	If not, why
1	CHIRPS (potential use) Rainfall	Climate Hazards Center	Climate Hazards Center	Spatial Data & 2D time varying	gridded	Yes	Yes	Yes	
2	GPM (potential Use) Rainfall	NASA	NASA	Spatial Data & 2D time varying	gridded	Yes	Yes	Yes	
3	ECMWF Reanalysis v5 (ERA5) Rainfall, Temperature, Radiation, humidity	ECMWF	Copernicus Climate Change Service (C3S) at ECMWF	Spatial Data & 2D time varying	gridded	Yes	Yes	Yes	
4	National Climate Data	Nat. Met Institutions	National met. institutions	Spatial Data & 2D time varying	gridded	Yes	Yes	Probably	Depends on national data sharing policies
5	Global Climate change prediction model results	Copernicus		Spatial Data & 2D time varying	gridded	Yes	Yes	Probably	Depends on national data sharing policies
6	Locally downscaled climate model results	Nat. Met Institutions/ PHISHES	National met. Institutions/ PHISHES partners	Spatial Data & 2D time varying	gridded	Yes	Yes	Probably	Depends on national data sharing policies
7	Other drivers to local hydrological models, applications of fertilizers and agrochemicals	Local authorities and stakeholders	Local authorities and stakeholders	Spatial Data & 2D time varying & Other data Type	gridded	Yes	Yes	No	Mandatory clause from data owners
8	Existing (calibrated) parameters sets for Hydrological models of case study areas	Local authorities and stakeholders	Local authorities and stakeholders	Spatial Data & 2D time varying & Other data Type	gridded	Yes	Yes	Probably Not	Pre project established model setups may be property of local authorities or stakeholders
9	Model results	PHISHES	PHISHES	Spatial Data & 2D time varying & Other data Type	gridded	Yes	Yes	Yes, with restrictions	Results should not be traceable to individuals or companies in the area



Table 2. PHISHES Impact of land use management on fate and transport of contaminant in soils (Work Package 2), Lead Partner: Bureau de Recherches Géologiques et Minières (BRGM)

	Data Description	Data Source	Origin of Data	Data Type	Data format	Metadata available	Keywords available	Can data be available	If not, why
1	Meteorological Forcing (Precipitation, Temperature)	Other sources (ECMWF)	DHI	2D time varying	NetCDF	Yes	No	Yes	
2	Observed Streamflow	Other sources (GRDC (WMO))	DHI	TS Data	CSV	No	No	No	GRDC often does not permit their data to be published other places
3	Simulated Runoff		DHI	2D time varying	Dfs2	Yes	Yes	Yes	
4	PFAS fate and transport monitored in Plurimetric Pilot	Other projects (PROMISCES)	BRGM	TS Data & Spatial Data	CSV	Yes	No	To be discussed	To be discussed with PROMISCES project
5	As(V) and As(III) fate and transport in old tailings before and after phyto-stabilisation	Other projects (oMIMo)	BRGM	TS Data & Spatial Data	CSV	Yes	No	Yes	
6	Pb and Zn fate and transport in old tailings before and after phyto-stabilisation	Other projects (PHOTOSELECT)	BRGM	TS Data & Spatial Data	CSV	Yes	No	Yes	
7	As(V) and As(III) fate and transport in contaminated soils	Other projects (Agris)	BRGM	TS Data & Spatial Data	CSV	Yes	No	Yes	



Table 3. PHISHES Test cases (Work Package 4.1), Lead Partner: CzechGlobe – Global Change Research Institute (CG)

	Data Description	Data Source	Origin of Data	Data Type	Data format	Metadata available	Keywords available	Can data be available	If not, why
1	Observed flow in streams	Other projects + generated in project	CG, PVL, VUMOP, CHMI	TS Data	csv, dfs0	Yes	No	Yes, with restrictions	Clear restriction from data owner, exact wording and terms in letter from PVL with SUBJECT: Consent to the use of data from the state enterprise Povodí Vltava for the EU HORIZONT project - PHISHES from 23.10.2024
2	Observed pollutant concentrations in streams	Other projects + generated in project	CG, VUMOP, PVL	TS Data	csv, dfs0	No	No	Yes, with restrictions	See 1.
3	Land management (crops type, harvesting, fertilizer etc.)	Other projects	VUMOP, PVL, individual farmers	Spatial Data	Shp+ CSV	No	No	No	Clear restriction from data owner.
4	Soil data (type, parameters, map)	Other projects	CG, VUMOP, CGS, PVL	Spatial Data	CSV+bit map	Yes	No	Yes, with restrictions	See 1.
5	Various climate data (T, P, Rad etc.)	Other projects	CG, CHMI, PVL	2D time varying	dfs2	Yes	No	Yes, with restrictions	See 1.
6	Simulated flow	Generated in project	CG	TS Data	csv, dfs0	Yes	No	Yes with restrictions	See 1.
7	Simulated WQ	Generated in project	CG	TS Data	csv, dfs0	Yes	No	Yes, with restrictions	See 1.
8	Climate data from CC scenarios	Other projects	CG	2D time varying	dfs2	Yes	No	Yes, with restrictions	See 1.
9	Hydrogeological parameters	Other projects	CGS, VUMOP, PVL	Spatial Data	dfs2	Yes	No	Yes, with restrictions	See 1.
10	Stream network + hydraulic parameters	Other projects	PVL, CG	Spatial Data	Mhydro	Yes	No	Yes with restrictions	See 1.

The main purpose of the data listed above is to serve as model input data as well as model calibration, validation and interpretation.

Abbreviations:

- 1) PVL - Povodí Vltavy (Vltava river authority), State Enterprise, <https://www.pvl.cz/en>
- 2) VUMOP - Research Institute for Soil and Water Conservation, <https://www.vumop.cz/en/>
- 3) CHMI - Czech Hydro-Meteorological Institute, <https://www.chmi.cz/?l=en>
- 4) CGS - Czech Geological Survey, <https://cgs.gov.cz/en>





Table 4. PHISHES Test cases (Work Package 4.2), Lead Partner: Warsaw University of Life Sciences (SGGW)

	Data Description	Data Source	Origin of Data	Data Type	Data format	Metadata available	Keywords available	Can data be available	If not, why
1	Streamflow data (discharge)	WP Project	Inst. Hydrol. & Meteor. Poland	TS Data	CSV, xlsx			Yes	
2	Air Temperature	WP Project	Inst. Hydrol. & Meteor. Poland	TS Data	CSV, xlsx			Yes	
3	Precipitation	WP Project	Inst. Hydrol. & Meteor. Poland	TS Data	CSV, xlsx			Yes	
4	Evaporation	WP Project	SGGW	TS Data	CSV, xlsx			Yes	
5	Groundwater levels (depth)	WP Project	SGGW	TS Data & Spatial Data	CSV, xlsx, shp			Yes	
6	Surface water levels	WP Project	Inst. Hydrol. & Meteor. Poland	TS Data	CSV, xlsx			Yes	
7	Surface water quality (nitrate, P, Cd, K)	Generated in project	SGGW & Polish Inspectorate of Env. Protection	TS Data	CSV, xlsx			Yes	
8	Soil water content	Generated in project	SGGW	TS Data	CSV, xlsx			Yes	
9	Soil, land use and topography maps	WP Project	WP Project (Elabor. MIKE SHE model)	Spatial Data	shp			Yes	



Table 5. PHISHES Test cases (Work Package 4.3), Lead Partner: Swedish Geotechnical Institute (SGI)

	Data Description	Data Source	Origin of Data	Data Type	Data format	Metadata available	Keywords available	Can data be available*	If not, why
1	Meteorological data (2023-2024)	Other projects & SMHI	SMHI	TS Data	CSV	Yes	No	Yes	
2	Precipitation (May 2025-)	Other projects	Geological Survey of Sweden	TS Data	CSV	Yes	No	Yes	
3	Selected PFAS concentrations Soil/Unsaturated/Saturated zone	Other projects	Geological Survey of Sweden	TS Data & 2D time varying	CSV	Yes	No	No	Data part of ongoing governmental assignment, and therefore will be published in accordance with assignment's time plan
4	Geological site-specific model	Other projects + partially Swedish Transport Adm.	Swedish Geotechnical Institute	Spatial Data	ASCII	Yes	No	No	Data possibly subject to secrecy requirements
5	Simulated saturated Flow	Other projects	Swedish Geotechnical Institute	Spatial Data & TS Data	Mfn	No	No	No	Data possibly subject to secrecy requirements
6	Selected moisture content UZ, measured	Other projects	Geological Survey of Sweden	2D time varying	CSV	No	No	No	Data part of ongoing governmental assignment within the Government of Sweden's Grant no. 1:4 for remediation and restoration of contaminated sites, and therefore will be published in accordance with assignment's time plan
7	Moisture content UZ, modelled	Other projects	Swedish Geotechnical Institute	2D time varying	CSV, mfn	No	No	No	See 6.
8	Groundwater chemistry data	Other projects	Geological Survey of Sweden	TS Data	CSV	Yes	No	No	See 6.
9	Simulated unsaturated flow	Other projects	Swedish Geotechnical Institute	2D time varying	CSV, mfn	No	No	No	See 6.
10	Selected estimated PFAS fluxes in unsaturated and saturated zone	Other projects	Swedish Geotechnical Institute	Spatial Data + 2D time varying	CSV, mts	No	No	No	See 6.

\* Data connected to the governmental assignment cannot be available now (and is marked as 'No') but will be available outside PHISHES following the assignments time plan so could be made available in the future.

The main purpose of the data listed above is to serve as model input data as well as model calibration, validation and interpretation.



As shown in the tables, an important part of the data in the different work packages will be used for modelling purposes, including input data and data needed for model calibration, validation and interpretation.

The types of data include hydro-meteorological and hydrometric data, data such as soil and vegetation to describe the physiography of the areas to be modelled as well as information on pollution, use of pesticides by the farmers, etc.

Much of the data originate from existing and/or previous projects and thereby this project is benefitting from previous project work. Similarly, we expect future projects will benefit from the work done in this project.

The data is primarily spatial and time series data, but other data types are also being used. The data format in most cases are commonly used data formats, which can easily be used by others.

The vast majority of the data will be available outside of the project. But there are a few cases where the data provider has specified restrictions on making the data publicly available. This includes e.g., pesticides data provided by the farmers, which has been provided on the condition that they are ensured anonymity. It is therefore important that this restriction is respected, not only for those who have provided the data, but also for those who are using the data. Some other datasets may be distributed with restrictions, e.g., only in an aggregated format which will maintain the anonymity. These restrictions will be further clarified in future versions of the DMP.

Apart from the data shown in the tables, initial information was also gathered on which work packages are expected to use, produce and benefit from the data. This will be further clarified and reported in later versions of the DMP.

## **3.2 Data Reuse**

Stakeholders that may have an interest in, and benefit from, reusing the project data have been initially identified as follows, and will be described in more detail in later versions of the DMP:

1. Decision-makers involved in assessing the effects of policies on soil health in terms of costs and benefits.
2. Agricultural agencies and advisers, who advise farmers as well as environmental agencies and institutions.
3. National environmental protection agencies.
4. Water boards and catchment managers, especially in the countries where the test cases will be conducted.
5. Institutions already engaged in this project and seeking to advance the science and technology beyond its current scope.
6. Other research institutions working with similar topics.



## 4. FAIR Principles

### 4.1 Data Findable

The PHISHES project datasets should be uniquely identified, where special and unique identifiers must be attached to each dataset that the project consortium intends to make open. This could be implemented through the usage of the Digital Object Identifier (DOI) standard, as it is the most common and widely adopted standard for the identification and citation of academic datasets and publications. This concerns everything related to datasets reference and name in addition to standards and metadata.

**All open datasets should be associated with a unique identifier such as a DOI.**

The DOI is an ISO (2022) standard that provides a permanent and unambiguous identification for digital objects. The DOI identifier is linked to metadata describing the object and other properties such as its location. One of the main advantages of the DOI is that the location of the object can change over time and the DOI identifier of the object can remain unchanged, and only the metadata associated with the object needs to change.

DOIs are assigned by registration agencies, managed by the International DOI Foundation (IDF). A DOI has a prefix and a suffix separated by a forward slash (/). The prefix identifies the Registration Agency, while the suffix is managed by the individual Registration Agency to uniquely identify the object within it. A Registration Agency may update the metadata of the DOIs it manages at any time, e.g., when it changes a URL.

**Each dataset should be associated with a unique name, abstract, keywords and metadata.**

Beside the unique identifier, each dataset ought to be associated with a unique name and an abstract, provided by the responsible owner or author of the dataset. The unique name should be descriptive in that it contains a summary of the specific nature of the dataset. In terms of abstract, a short description summarising the aim and main features of the dataset is recommended. When possible, it is encouraged to refer to the name of the project (PHISHES) and the granting authority.

Once stored, the data needs to be findable by e.g., search engines. The primary mechanism for this from the data creator's perspective is to add metadata at publishing time to enable searching via e.g., keywords or tags. Metadata is a set of data that describes and gives information about a certain dataset. Different metadata standards have been developed and, among these different available standards, Dublin Core is recommended as it is a well-established standard for basic domain agnostic metadata describing resources of many different types. Dublin Core provides a set of 15 basic elements suitable for describing a resource and its attributes are shown in Table 6.

For the data produced by PHISHES, the dataset owner will be responsible for providing a unique name, abstract, metadata and keywords for each dataset. A template for providing these details could be established in future versions of the DMP, and partners would have access to this template through the project internal PHISHES Project Collaborative Space (the SharePoint site). For the generally available European or Global datasets, such metadata will be listed to the extent they are available from the provider.



Table 6. Metadata standard elements for describing a resource and its attributes

Label	Definition
Contributor	An entity responsible for making contributions to the resource.
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
Creator	An entity primarily responsible for making the resource.
Date	A point or period associated with an event in the lifecycle of the resource.
Description	An account of the resource.
Format	The file format, physical medium, or dimensions of the resource.
Identifier	An unambiguous reference to the resource within a given context.
Language	A language of the resource.
Publisher	An entity responsible for making the resource available.
Relation	A related resource.
Rights	Information about rights held in and over the resource.
Source	A related resource from which the described resource is derived.
Title	A name given to the resource.
Type	The nature or genre of the resource.

## 4.2 Data Accessible

The PHISHES project data policy should promote the Data Sharing principle. The PHISHES project will follow the principle ‘as open as possible as closed as necessary’, unless providing open access would in particular:

- Be against the beneficiary’s legitimate interests, including regarding commercial exploitation, or
- Be contrary to any other constraints, in particular the EU competitive interests or the beneficiary’s obligations under this Agreement; if open access is not provided (to some or all data), this must be justified in the DMP.

The PHISHES project input/output datasets, intermediate data products, and final information products will be stored in a **secure project data repository**. The datasets will also be made available to the project partners by the **PHISHES Digital Platform (PDP)** of the project using interoperable data services. As a general approach, PHISHES makes the interoperability of datasets and data collection one of the crucial features and constraints for the implementation of the PDP. Later versions of the DMP will specify the details of the PHISHES dataset repository.



PHISHES will promote FAIR principles, therefore datasets and products generated by the PHISHES project and approved by the data owners will be made accessible outside of the project through an **Open Repository** (such as Zenodo) with a unique identifier and detailed metadata for Open Search and Open Access. The datasets and products will be made publicly available once approval has been given by the data owner. The PHISHES **project website** will point to this open access tool.

This will ensure open access under the latest available version of the Creative Commons Attribution International License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a license with equivalent rights. Metadata of datasets should be open under a Creative Commons Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles (in particular machine-actionable). All EU requirements detailed in the project GA are thoroughly articulated in the PHISHES Project Handbook (Deliverable 7.1).

It is acknowledged that datasets produced in other projects and used in the PHISHES project may be published elsewhere. The PHISHES project website could point to these locations.

As stated in the Consortium Agreement (CA), there are some types of data, know-how, and information that are subject to specific limitations. These limitations fall under two broad categories: restrictions for implementation and restrictions for exploitation. Broadly, this describes the limitations within and outside of the project. These restrictions include, for example, that unpublished data should not be shared or used by the project beneficiaries beyond the project activities (only to be used during project implementation). Any changes to the agreed upon restrictions to sharing data must be approved by the General Assembly (GLA).

Data which will not be made openly accessible will include data which contains personally identifiable information, e.g., individual evaluation forms and data underlying deliverables that are covered by confidentiality. The personal data processed in the project are not made publicly accessible but are kept closed and inaccessible to third parties.

**Data owners must indicate whether a dataset can be shared (with or without restrictions) or if it cannot be shared, along with the reason for its restriction.** This process, including details for access to datasets that can only be accessed with restrictions, will be detailed in later versions of the DMP. This could include, for example, a tag assigned to each dataset in the secure project data registry stating if the data is 'Open', 'Open with restrictions' or 'Restricted', with an accompanying specification of the restrictions and reasons.

In cases where the datasets cannot be publicly shared outside of the project, the reason will be indicated (e.g., ethical issue, rules of personal data, intellectual property, commercial datasets, privacy related, security related). Metadata for restricted datasets could be exposed in the project website, with prior approval from the data owner, to specify whom to contact (e.g., the data owner) to acquire possible access to the data.

**Before external publication or dissemination of any kind, including scientific dissemination, approval from stakeholders / data owners must be secured.** This process is to go through the Work Package Leader and, in particular for data in regard to the PHISHES test cases, the process is to go through the test case leaders. More information on this process will be provided in future versions of the DMP.

For detailed information on dissemination of results, please see the PHISHES Project Handbook (Deliverable 7.1) and the Communication and Dissemination Strategy (Deliverable 6.2).

Timelines for depositing data in the Open Repository will be set out in future versions of the DMP. Data made accessible outside of the project through the Open Repository (such as Zenodo) will remain available and findable for as long as the Open Repository is maintained.

### 4.3 Data Interoperable

Interoperable data means that the datasets are available in a standard data format and exposed by a standard data access service. Data can be made available in fact in many different formats implementing different information models. The heterogeneity of these models reduces the level of



interoperability that can be achieved. Key to interoperable data exchange are standardised interfaces.

In general, the reused and generated datasets that will be available to the project partners within the PDP should be accessible through a set of standard APIs (such as the Open Geospatial Consortium - OGC - services) in such a way that they can be accessible by all the targeted users within the project.

Open datasets made available through the Open Repository (once approved by the data owners) should be, where possible, in a standard data format such as JSON, XML, CSV, NetCDF or similar.

Further details on standardisation of data formats, access services and interfaces (e.g., in the project PDP and project website) will be provided in later versions of the DMP.

#### **4.4 Data Reusable**

Open data produced by the PHISHES project and made available through the Open Repository (e.g., Zenodo), when approved by the data owner, will be freely available in the public domain to support data re-use.

Information will be provided via the project website or the Open Repository about any software or tools needed to facilitate data re-use, and further details of how the information will be provided will be specified in future versions of the DMP.

PHISHES beneficiaries acting as data owners are responsible for quality assurance of their datasets shared with project partners via the secure data repository and PDP, and of datasets openly available on the Open Repository.



## 5. Data Security

The PHISHES project deals with a large amount of data, some of which is considered as personal data and some of which is subject to specific limitations as set out in the Consortium Agreement (CA) and also in the DMP. The consortium is therefore aware of the need for securing data analysis and datasets that will be generated within the project, and, to this end, data security is considered a priority in the PHISHES project.

Special attention will be given to the security of new data being generated within the project using the different appropriate methods with respect to the types of datasets and metadata mentioned in the previous sections. Moreover, the project consortium prioritises the protection of personal data that will be collected in the project where measures will be taken, e.g., the anonymisation/pseudonymisation/-encryption methods to protect personal data.

The PHISHES project data files will be managed, processed, and stored in a secure data repository, and protected by controlled access to digital files with encryption and/or password protection. Possible anonymisation of data will be implemented as early in the data processing pathway as possible, preferably before the data is stored in the repository.

Participating partners involved in data collection and processing will be individually responsible for maintaining and observing agreed data security protocols, including for any working and development versions of locally stored data e.g., on local hardware or servers.

### 5.1 Data preservation and processing

Within the PHISHES project the responsibility to maintain and preserve the datasets will fall under the hands of the dataset owner and author when possible. In case, during or after the project, the capability of the project partners to maintain the dataset changes, and they are no longer able to maintain the dataset, a dataset management transfer, including the ownership, will be considered, and evaluated based on the relevance, the value, and the related costs.

The ownership, authorship, responsibility and any associated licenses of each dataset will be clearly described, reported, and continuously updated in later versions of the DMP when other related deliverables get developed.

The PHISHES project data files will be stored in a secure project data repository which will be backed up on a regular basis in the form of e.g., database dumps. Each partner must keep a backup of the data the partner stores in the repository. The database snapshots should be archived using a suitable data repository that allows issuing metadata with the data. The metadata will contain all relevant licensing, versioning and related publications or previous data sets. Any upgraded or newer versions of data will be issued with metadata that includes information on any previous versions. Within the repository, older versions will be marked, and users will be directed to the updated information.

All the processing tools for the input/output data will ensure the application of the FAIR principles and EU requirements. The PHISHES consortium fully understands the need for the design of data processing tools that adhere to the FAIR principles as articulated in the section above, and they will be reported in the updated versions of the DMP.





## 6. Allocation of Resources

The project has allocated a budget for the management of the project data.

Each data provider is responsible for providing clear and accurate information regarding their data. Work package “Project coordination and management” (Work Package 7) is responsible for overall management of data in the project.

More details will be provided in later versions of the DMP as needed.



## 7. Privacy and Ethical Issues

Within the PHISHES project all matters related to privacy and data protection will be conducted in accordance with (EU) 2016/679 (GDPR) (2016) and Regulation (EU) 2018/1725 (EUDPR) (2018). Several measures must be taken and will continuously be improved throughout the life of the project, to ensure all privacy and data protection policies get implemented by all the partners within the project.

All topics related to the processing of sensitive information and EU classified information will be identified and agreed on with all the project partners.

The project partners involved in data collection will be responsible for maintaining and observing agreed data collection and security protocols including the preservation of the anonymity of participants.

All project partners have agreed in the Consortium Agreement to cooperate, where necessary, to enable one another to fulfil legal obligations arising under applicable data protection laws within the scope of the performance and administration of the Project and of the CA. Furthermore, all project partners have agreed in the CA to conclude, where necessary, a separate data processing, data sharing and/or joint controller agreement before any data processing or data sharing takes place.

Personal data shall always be treated as Confidential Information and shall be protected with an adequate level of safety and confidentiality, subject to any applicable legal, regulatory or contractual requirements. Informed consent should be sought from each participant for the long-term preservation and sharing of anonymised respondent data with the respondents right to withdraw clearly outlined.

Concretely, the PHISHES project deals with the following aspects that can be considered as Ethical Issues:

- The collecting, processing, and storing of personal data.
- The involvement of humans in data collection activities.
- The requirement of natural persons.
- Health and safety of the people working in the project and beneficiaries.



## 8. Scientific Publications

All beneficiaries in the PHISHES project must ensure open access to peer-reviewed scientific publications relating to their results.

To automate the process of reporting scientific publications and related research data in OpenAIRE, the publication should be deposited in an OpenAIRE-compliant repository such as Zenodo, either by the authors of the publication (green open access) or by a scientific publisher (gold open access). The details of the repository will be provided in later versions of the DPM.

Additional forms of disseminating open access papers, including academic social network sites such as ResearchGate will be considered.

In addition, all publications should be available in the PDP where possible.

For details on open science and access to publications in the PHISHES, please see the PHISHES Project Handbook (Deliverable 7.1).

### 8.1 Green open access

Green open access or self-archiving means that the published article or the final peer-reviewed manuscript is archived by the researcher itself in an online repository. The journal must grant the researcher the permission to self-archive the final peer-reviewed article, at the latest, 12 months after publication.

### 8.2 Gold open access

Gold open access means that the publication is available by the scientific publisher as open access. Some journals require an author-processing fee for publishing open access. Some publishers allow the researcher to deposit a copy of the article in a repository, sometimes with an embargo period.



## 9. Conclusion

In this version of the Data Management Plan, a preliminary overview of the data that will be processed and generated in the PHISHES project is provided. The input and output datasets have been identified, in addition to highlighting the FAIR principles, ethical issues, open science practices, and the process to manage scientific publications. This DMP is a live document and will be updated regularly throughout the life of the project, where it will be possible to elaborate further on different aspects, i.e., input/output datasets and FAIR principles.



## 10. References

Dublin Core, <https://www.dublincore.org/about/>

European Commission: European Innovation Council and SMEs Executive Agency, European IP helpdesk – Your guide to open science in Horizon Europe, Publications Office of the European Union, 2024, <https://data.europa.eu/doi/10.2826/943044>

European Open Science Cloud (EOSC) <https://open-science-cloud.ec.europa.eu/>

Horizon Europe Open Access Guidelines, 2017, [https://enspire.science/wp-content/uploads/2019/12/h2020-hi-oa-pilot-guide\\_en.pdf](https://enspire.science/wp-content/uploads/2019/12/h2020-hi-oa-pilot-guide_en.pdf)

Horizon Europe Programme, Data Management Plan Template, (HE) (V1.1 – 01.04.2022), 2022, <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/reference-documents?programmePeriod=2021-2027>

ISO, 26324, 2022, <https://www.iso.org/standard/81599.html>

Open Research Europe (ORE) <https://open-research-europe.ec.europa.eu/>

Regulation EU 2016/679 of the European Parliament, 2016, <http://data.europa.eu/eli/reg/2016/679/2016-05-04>

Regulation EU 2018/1725 of the European Parliament, 2018 <http://data.europa.eu/eli/reg/2018/1725/oj>

Regulation (EU) 2021/695 of the European Parliament, 2021, <http://data.europa.eu/eli/reg/2021/695/2024-03-01>

ResearchGate, <https://www.researchgate.net/>

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018, 2016, <https://doi.org/10.1038/sdata.2016.18>



**PHISHES**

Physically-based Integrated  
Soil Health Simulation platform

*Predicting soil health to protect ecosystems*

### For More Information

- Visit PHISHES website:  
[www.phishes-project.eu](http://www.phishes-project.eu)
- Follow PHISHES on LinkedIn:  
[www.linkedin.com/company/phishes-project](http://www.linkedin.com/company/phishes-project)



**Funded by  
the European Union**

*Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.*