

D1.3 Data Management Plan

Authors: Ioannis N. Daliakopoulos, Thrassyvoulos Manios

REACT4MED Project (Grant Agreement No. 2122) is funded by PRIMA - Partnership for Research and Innovation in the Mediterranean Area, a Programme supported by Horizon 2020.



PRIMA
PARTNERSHIP FOR RESEARCH AND INNOVATION
IN THE MEDITERRANEAN AREA





This page was intentionally left blank

Technical References

Project Acronym	REACT4MED
Project title	Inclusive Outscaling of Agro-ecosystem REstoration ACTions for the MEDiterranean
Project coordinator	HMU
Project Duration	36 months
Deliverable No.	D1.3
Deliverable title	Data Management Plan
Lead partner	HMU
Contributing partners	
Author(s)	Thrassyvoulos Manios, Ioannis Daliakopoulos
Editor(s)	Thrassyvoulos Manios
Type	Text
Format	MS-Word
Language	EN-GB
Creation date	2022.10.31
Version number	1.0
Version date	
Last modified by	
Due date	2022.10.31
Actual delivery date	
Rights	Copyright © 2022, REACT4MED Consortium
Dissemination level	<input type="checkbox"/> CO (confidential, only for members of the consortium)
	<input checked="" type="checkbox"/> PU (public)
	<input type="checkbox"/> PP (restricted to other programme participants)
	<input type="checkbox"/> RE (restricted to a group specified by the consortium)
	When restricted, access granted to:
Nature	<input checked="" type="checkbox"/> R (report)
	<input type="checkbox"/> P (prototype)
	<input type="checkbox"/> D (demonstrator)
	<input type="checkbox"/> O (other)

Revision history			
Version	Date	Modified by	Comments
1.0	2022.10.31	ID_HMU	First version of the DMP

Copyright © 2022, REACT4MED Consortium

The REACT4MED Consortium grants third parties the right to use and distribute all or parts of this document, provided that the REACT4MED Project and the document are properly referenced.

THIS DOCUMENT IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS DOCUMENT, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Table of Contents

Executive summary	iv
1 Introduction.....	1
2 Data Summary	2
2.1 Purpose of Data Collection	2
2.2 Types and Formats of Collected/Generated Data.....	2
2.2.1 Background data	2
2.2.2 Foreground data	2
2.3 Standards and metadata	3
2.4 Data sharing	3
2.4.1 Sharing through the REACT4MEDhub data sharing platform.....	3
2.4.2 Sharing through other cloud storage services	3
2.5 Archiving and preservation (including storage and backup)	3
3 FAIR data.....	3
3.1 Findable Data	3
3.2 Accessible	4
3.3 Interoperable	4
3.4 Re-use Optimization.....	4
3.5 Version Control.....	4
4 Naming Conventions.....	5
References.....	5

Executive summary

REACT4MED will generate a considerable amount of intermediate (WP4) and final data (WP5) as well as transferable knowledge (WP6). This information will be created both top-down (from WP to demo sites) and bottom-up (from stakeholders and Pilot Areas to WPs) and REACT4MED will involve many people with diverse backgrounds, and the potential for project activities to become fragmented. Furthermore, while REACT4MED aims to make as many outputs as possible freely available and accessible, disclosure of knowledge can be sensitive, for example in the cases where original ideas revealed to early in their development may be exploited by outsiders. To effectively safeguard those whose knowledge needs protection (both within and outside the consortium) and protect Intellectual Property Rights (IPR), a dedicated plan including specific protocols will be developed and implemented in accordance with the principles of the Open Research Data Pilot. Also, in accordance with the Open Research data Pilot, partners in the REACT4MED consortium will share their research results within the project and/or publicly with minimum delay using the most appropriate channels (i.e., for peer-reviewed scientific papers, (gold) open access publishing will be promoted, which is this reflected by an adequate budget allocated to all partners in WP7). The IPR protocol will be drafted within the Dissemination and Communication Strategy (Task 7.2) and agreed at Scientific Board level and all partners will be asked to abide by its terms and conditions.

REACT4MED results will integrate a variety of data ranging from static spatial layers and short range environmental timeseries to spatiotemporal climate model datasets. Besides scale diversity and storage requirements, use and distribution license considerations should also be made. The flow of data to be collected, processed and/or generated by REACT4MED throughout its lifecycle will be described in the Data Management Plan (DMP). The DMP will be developed early in the project (month 6) and set out mechanisms and procedures to render REACT4MED results findable, accessible, interoperable, and re-usable (FAIR). Additionally, most results and communication or dissemination products will also be made publicly available, and stored REACT4MED on the project's own website and online information system as soon as is practical. It is envisaged that REACT4MED will provide a repository for all deliverables and outputs acquired during the project. Data and other information required to support and allow reproduction of project results included in the REACT4MED dissemination and exploitation will be shared and readily available on public (open access) online repositories (e.g., ZENODO). Costs for hosting the data and the REACT4MEDHub will be covered from the PRIMA Grant and are guaranteed to be available after the lifetime of REACT4MED according to the DMP for the sustainability of the project output.

In consultation with Pilot Area leaders where most REACT4MED data will be collected as well as WP leaders where most data will be processed and generated, a Data Management Plan (DMP) will be compiled. Each product will be annotated with a list of attributes including Responsible Investigator(s), Data Product Description, Product Format(s), Anticipated Data Volume, Intended Repository, Release Timeline, License for Reuse, Duration of Data Availability. The DMP will also consider other products such as software where relevant information (e.g., dependencies and programming language) will be stored. The process of annotation will be streamlined using free tools (e.g., ezDMP). Unless restricted by license agreements, data collected, and the outputs of the research activities will be shared and readily available on public (open access) online repositories (e.g., ZENODO). To put Pilot Area data in context, relevant metadata from the Pilot Areas will also be included (see Pilot Area descriptions). Both the project public deliverables as well as the scientific papers will be available through open access channels. A DMP report including the workflow and initial data to be collected will be compiled early in the Project (M6) and will be updated throughout the project.

1 Introduction

REACT4MED will generate a considerable amount of intermediate (WP4) and final data (WP5) as well as transferable knowledge (WP6). This information will be created both top-down (from WP to demo sites) and bottom-up (from stakeholders and Pilot Areas to WPs) and REACT4MED will involve many people with diverse backgrounds, and the potential for project activities to become fragmented. Furthermore, while REACT4MED aims to make as many outputs as possible freely available and accessible, disclosure of knowledge can be sensitive, for example in the cases where original ideas revealed to early in their development may be exploited by outsiders. To effectively safeguard those whose knowledge needs protection (both within and outside the consortium) and protect Intellectual Property Rights (IPR), a dedicated plan including specific protocols will be developed and implemented in accordance with the principles of the Open Research Data Pilot. Also, in accordance with the Open Research data Pilot, partners in the REACT4MED consortium will share their research results within the project and/or publicly with minimum delay using the most appropriate channels (i.e., for peer-reviewed scientific papers, (gold) open access publishing will be promoted, which is this reflected by an adequate budget allocated to all partners in WP7). The IPR protocol will be drafted within the Dissemination and Communication Strategy (Task 7.2) and agreed at Scientific Board level and all partners will be asked to abide by its terms and conditions.

REACT4MED results will integrate a variety of data ranging from static spatial layers and short range environmental timeseries to spatiotemporal climate model datasets. Besides scale diversity and storage requirements, use and distribution license considerations should also be made. The flow of data to be collected, processed and/or generated by REACT4MED throughout its lifecycle will be described in the Data Management Plan (DMP). The DMP will be developed early in the project (month 6) and set out mechanisms and procedures to render REACT4MED results findable, accessible, interoperable, and re-usable (FAIR). Additionally, most results and communication or dissemination products will also be made publicly available, and stored REACT4MED on the project's own website and online information system as soon as is practical. It is envisaged that REACT4MED will provide a repository for all deliverables and outputs acquired during the project. Data and other information required to support and allow reproduction of project results included in the REACT4MED dissemination and exploitation will be shared and readily available on public (open access) online repositories (e.g., ZENODO). Costs for hosting the data and the REACT4MEDHub will be covered from the PRIMA Grant and are guaranteed to be available after the lifetime of REACT4MED according to the DMP for the sustainability of the project output.

In consultation with Pilot Area leaders where most REACT4MED data will be collected as well as WP leaders where most data will be processed and generated, a Data Management Plan (DMP) will be compiled. Each product will be annotated with a list of attributes including Responsible Investigator(s), Data Product Description, Product Format(s), Anticipated Data Volume, Intended Repository, Release Timeline, License for Reuse, Duration of Data Availability. The DMP will also consider other products such as software where relevant information (e.g., dependencies and programming language) will be stored. The process of annotation will be streamlined using free tools (e.g., ezDMP). Unless restricted by license agreements, data collected, and the outputs of the research activities will be shared and readily available on public (open access) online repositories (e.g., ZENODO). To put Pilot Area data in context, relevant metadata from the Pilot Areas will also be included (see Pilot Area descriptions). Both the project public deliverables as well as the scientific papers will be available through open access channels. A DMP report including the workflow and initial data to be collected will be compiled early in the Project (M6) and will be updated throughout the project.

2 Data Summary

2.1 Purpose of Data Collection

The DMP manages the outputs generated from all seven work packages during the course of REACT4MED. The project is structured in seven thematic work packages WP1-7. WP2, 3, 4, and 6 (baseline data, living labs engagement, modelling, policy) is dedicated to the setup of the REACT4MED ecosystem restoration and WP5 tests and implements the ecosystem of innovation in eight pilot areas across the Mediterranean region. Exploitation and dissemination activities in WP7 and scientific coordination and management in WP1 complement the four thematic work packages. Collecting data from all eight pilot areas establishes a comprehensive approach to land restoration that produces robust estimates of effectiveness, develops and tests approaches.

2.2 Types and Formats of Collected/Generated Data

2.2.1 Background data

According to the Grant Agreement (Article 16.1), background is defined as “data, know-how or information (...) that is (...) needed to implement the Action or exploit the results”. Because of this need, Access Rights have to be granted in principle, but Parties must identify and agree amongst them on the Background for the Project. In this context, Table 1 lists the Background data, along with the contributing party and specific restrictions and/or conditions for implementation (see Article 16.4 Grant Agreement, and its Annex 5, Section “Access rights to results and background and results for implementing the Action”), and exploitation (see Article 16.4 Grant Agreement and its Annex 5, Section “Access rights to results and background”, sub-section “Access rights for exploiting the results”).

Table 1: Background data needed for the implementation and exploitation of REACT4MED.

Contributing Party	Describe Background	Specific restrictions and/or conditions for implementation	Specific restrictions and/or conditions for Exploitation

2.2.2 Foreground data

During the project, REACT4MED will collect data of various types (e.g., Agronomic, Meteorologic, Climatic, Socioeconomic, Land use). Table 2 lists (indicatively) the data to be produced by the consortium along with their data types (e.g., variable or set of variables), data origin and data provider, data format (i.e., file type), WP for which data use is intended or requested, expected data size, and restrictions or copyrights.

Table 2: Foreground data needed for the implementation and exploitation of REACT4MED.

Category	Data Type	Data Origin/ Data Provider	Data Format	Data Use (WP)	Expected Data Size	Restrictions / copyrights
Agronomy	Yield	HMU	.xlsx	4	KB (kilobytes)	PU
Meteorologic	Data from the meteorological station of HMU	HMU	.csv	4	MB	TBD

2.3 Standards and metadata

2.4 Data sharing

2.4.1 Sharing through the REACT4MEDhub data sharing platform

The consortium will be sharing, the documents produced, via MS Teams as shown below

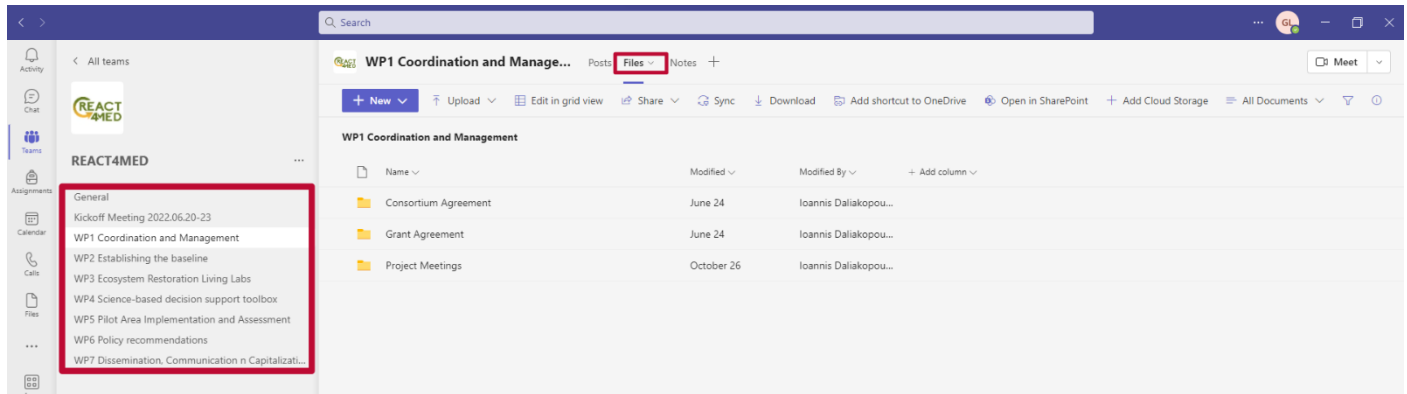


Figure 1: Files shared between partners

2.4.2 Sharing through other cloud storage services

In some cases, data will be of significant size or for very specific use and not convenient to store and share on the REACT4MED data sharing platform (Microsoft Teams). In these cases, rather than including this data, links to this data can be provided in word documents that can also function as README files (see Section 3.3 – Reuse optimization).

2.5 Archiving and preservation (including storage and backup)

Microsoft Teams keeps automated backups of all Microsoft Office documents uploaded. Screenshots in Figure 2 provided below show the steps finding previous revisions of the same document.

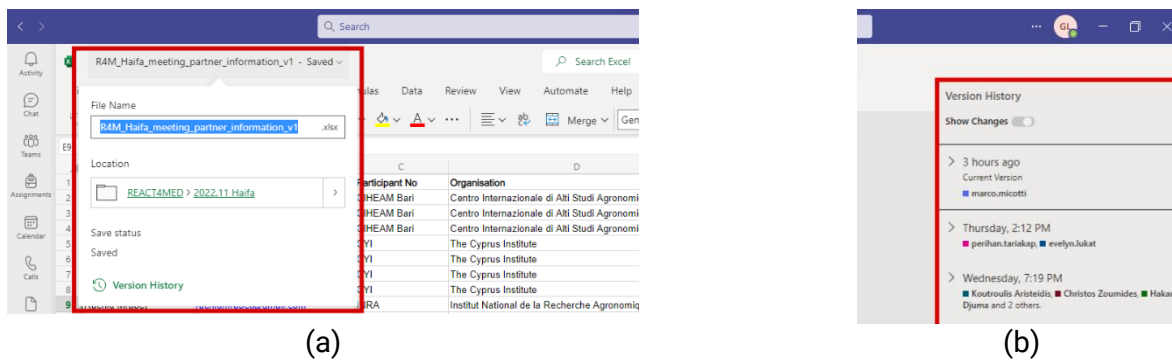


Figure 2: Access of document history through Microsoft Teams.

Apart from this history, backups will be kept in physical mediums every 6 months by HMU.

3 FAIR data

3.1 Findable Data

For the data to be findable, every metadata record requires a unique identifier to provide certainty to the identity of the record, and to lay out a primary key for linkages (Koers et al., 2020). The identifier must remain

distinct and invariant, irrespective of where the metadata record is stored. This allows linkages to a metadata record to persist for long-term data storage and preservation.

Publishing the REACT4MED outputs in an open repository service, such as GitHub, makes them citable by archiving the repository in a data-archiving tool, such as Zenodo that assigns a DOI (Sicilia et al., 2017), which is the backbone of the academic reference, to each record. The project Consortium may consider this option for the data management plan. GitHub is one of the wide-reaching and most popular repository hosting services. Those repositories can be archived using Zenodo, which ensures that all metadata required for the identification of the repositories are filled before the final public release. Operated by CERN, Zenodo aggregates EU funded research output from thousands of repositories available worldwide, links them to grants from EU Commission, and makes them available by indexing them via the OpenAIRE portal, free of charge (Commission et al., 2019). REACT4MED partners will also make publications and research outputs available in selected renowned data portals, first quartile (Q1) journals, and widely distinguished platforms among the climate and water community (*OpenAIRE*, 2022). This could include sharing databases via broadly disseminated portals such as EEA data (WISE), Climate Adapt portal, World Data Center, to name a few.

3.2 Accessible

Datasets tend to degrade or disappear over time because there is a cost to maintaining an online presence for data resources. When this happens, links become invalid and users waste time hunting for data that might no longer be there. Storing the metadata generally is much easier and cheaper. Thus, all data provided should be accompanied by a read me file (see paragraph 3.4).

3.3 Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing. Data should thus be in an open file format making them accessible to proprietary, free, or open-source software.

3.4 Re-use Optimization

As part of the publication process, data-archiving services such as Zenodo provide the association of search keywords with datasets using the menu to the right of Zenodo publication page. This allows search engines to identify and index related files automatically, thus optimizing wider possibilities for re-use. Suggested keywords may include, but are not limited to climate change; water scarcity; adaptation; Mediterranean; remote sensing; stakeholder; socio-hydrology; water management; water accounting; innovative irrigation technologies and practices; adaptation and mitigation strategies.

Data that is not self-explanatory should be accompanied by a README text file (e.g., README.md or README.txt) that provides relevant information (e.g., data overview, documentation, installation, how to cite, contributors, copyright, acknowledgement).

3.5 Version Control

Data files should be accompanied by a revision history (i.e., in the README file). An indicative revision history table is shown in Table 3.

Table 3: Revision history data kept for version control.

Revision history			
Version	Date	Modified by	Comments
1.0			

4 Naming Conventions

Following a consistent and precise naming convention facilitates the process of dataset access and retrieval for the future scientific and broader community (*OpenAIRE*, 2022).

R4M encourages the use of a standard naming convention given to all its public domain documents as follows:

“R4M_YYYY_WPX#_DOC#_DOCKEYWORD_”

- “R4M” stands for REACT4MED.
- “_” is a underscore that indicates a separator between elements.
- “YYYY” stands for the 4-digit year.
- “WPX#” stands for the work package under which the data lies.
- “DOC#” stands for the document number assigned to each file.
- “DOCKEYWORD” indicates a keyword associated with the file that identifies it further.

This naming convention may be revisited throughout the course of the project based on generated outputs.

References

- Commission, E., Innovation, D.-G. for R. and, Wareham, J., & Pujol Priego, L. (2019). *Zenodo : open science monitor case study*. Publications Office. <https://doi.org/doi/10.2777/298228>
- Koers, H., Bangert, D., Hermans, E., van Horik, R., de Jong, M., & Mokrane, M. (2020). Recommendations for Services in a FAIR Data Ecosystem. *Patterns*, 1(5), 100058. <https://doi.org/10.1016/J.PATTER.2020.100058>
- OpenAIRE*. (2022). <https://www.openaire.eu/>
- Sicilia, M. A., García-Barriocanal, E., & Sánchez-Alonso, S. (2017). Community Curation in Open Dataset Repositories: Insights from Zenodo. *Procedia Computer Science*, 106, 54–60. <https://doi.org/10.1016/j.procs.2017.03.009>