D1.2 "INITIAL DATA MANAGEMENT PLAN (DMP)"

NOVA

Funded by
the European Union
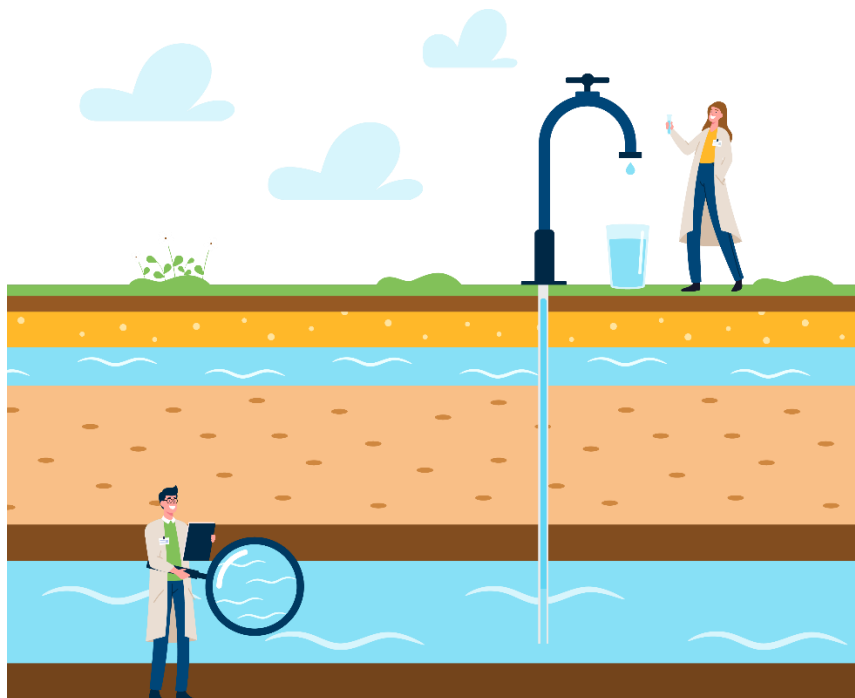
GRANT AGREEMENT NUMBER: 101082048

PROJECT ACRONYM: MAR2PROTECT

PROJECT TITLE: "Preventing Groundwater Contamination Related to Global and Climate Change through a Holistic Approach Based on Managed Aquifer Recharge"

PROJECT Duration: 1st December 2022 - 30th November 2026 (48 months)

WEBSITE: https://mar2protect.eu/

ABBREVIATION / ACRONYM:

Abbreviation / Acronym	Description
DMP	Data Management Plan
EU	European Union
FAIR	Findability, accessibility, interoperability, and reusability
WP	Work Package
RAID	Risks, Assumptions, Issues and Dependencies
DOI	Digital Object Identifier
IP	Intellectual Property
PT	Portugal
SciFinder	CAS - Division of the American Chemical Society
OpenAIRE	European Open Science Infrastructure, for open scholarly and scientific communication
HTTPS	Hypertext Transfer Protocol Secure
DPO	Data Protection Officer
GDPR	General Data Protection Regulation – European Union
GW	Groundwater
GC	Global change
CC	Climate change
MAR	Managed Aquifer Recharge
M-AI-R DSS	New-generation MAR decision support system based on AI-based techniques
WW	Wastewater
SW	Surface water
RCPs	Representative Concentration Pathways

Funded by
the European Union

LCA	Environmental assessment: Life Cycle Assessment
LCC	Economic assessment: Life Cycle Cost
S-LCA	Social Life Cycle Assessment

EXECUTIVE SUMMARY

The following document is Deliverable 1.2 - Data Management Plan (DMP) of the MAR2PROTECT Project, funded by the European Union's Horizon Europe research and innovation programme under grant agreement Number **101082048**.

This document is the first version of the DMP, consisting of preliminary information regarding the type and format of data that will be collected and generated, its origin, data utility and how MAR2PROTECT's research data will be findable, accessible, interoperable, and reusable (i.e., FAIR).

The purpose of the DMP is to provide the members of the consortium with an analysis of the main elements of the data management policy regarding all the datasets generated by the project.

Deliverable Number	WP / T
D1.2	WP1 / T1.3

Lead Beneficiary	Deliverable Author (S)
NOVA	Ana B. Pereiro

Beneficiaries	Deliverable Co-Author (S)
UNIBO	Dario Frascari
FEUGA	Tamara Rodríguez & Ana Vila Ferreiro

Planned Delivery Date	Actual Delivery Date
31.05.2023	02.05.2023

Type of deliverable	PU	Public, fully open, e.g. web (Deliverables flagged as the public will be automatically published on the CORDIS project's page))	**X**
	SEN	Sensitive, limited under the conditions of the Grant Agreement	

REVISION HISTORY

Version	Date	Author	Document history/approvals
01	28/11/2022	NOVA	The initial draft for circulating to project partners
02	02/12/2022	FEUGA	1st revision of FEUGA
03	05/12/2022	UNIBO	1st revision of UNIBO
04	31/01/2023	All partners	1st revision of all partners
05	02/05/2023	NOVA	The final draft for circulating to project partners
06	24/05/2023	NOVA	Added comments and fixes from partners. Added an Excel file for dataset register

**Funded by
the European Union**

Index

**Funded by
the European Union**

Figure Index

Table Index

1 Data Summary

1.1 Purpose of the data collection/generation

The purpose of data generation and collection in the MAR2PROTECT project is to obtain quantitative and qualitative data to successfully achieve MAR2PROTECT´s objectives, namely:

- the prevention of groundwater (GW) contamination from the impacts of global change (GC) and climate change (CC) based on a new-generation Managed Aquifer Recharge (MAR).
- the creation of a new development of a decision support tool (M-AI-R DSS) that will incorporate technological and societal engagement information to improve the groundwater quality and quantity.

The data generation and collection will comply with the EU ethics and legal requirements as well as national ethics and legal requirements.

1.2 Data types and formats

MAR2PROTECT project will both generate and reuse different types of data as a result of the different analyses that will be performed in the project: pollutant concentration in real-time from wastewater/surface water(WW/SW); data on pollutant removal from wastewater/surface water; aquifer characterization data; meteo data; CC/GC scenarios (RCPs, sea level, land use & cover, population growth); MAR yield predictions; risk indicators GIS maps; data provided to, collected from and about citizens and other stakeholders; environmental assessments (Life Cycle Assessment, LCA), economic assessments (Life Cycle Cost, LCC), and Social Life Cycle Assessment (S-LCA) of technologies; market entry assessment to identify the best route to market; a preliminary business model for most promising technologies; patentability analysis; state of the art; a proposal of optimal IP protection; exploitation strategy of project results; etc...

Each beneficiary must ensure open access (free of charge online access for any user) to all peer-reviewed scientific publications relating to its results.

Moreover, the beneficiary must aim to deposit at the same time the research data needed to validate the results presented in the deposited scientific publications.

MAR2PROTECT project data will be stored in the next type of files:

- Any kind of elaborated document in PDF format (**DOC**).
- ASCII Data files in CSV format with a related metadata document (**ASCII**).
- Structured data files with their metadata generated by proprietary applications (**SDF**), including GIS files that will be supplied in an open format with its metadata.

**Funded by
the European Union**

Table 1. *Types of documents and data and formats of data generated and collected during the project.*

Data Format	Data description	Work Package
ASCII	Characterization of (bio)materials	WP2
	Pollutant removal capacities from WW/SW	WP2
	Monitoring data: pollutant concentration in real-time from WW/SW	WP3
	Meteo data	WP4
	CC/GC scenarios (RCPs, sea level, land use & cover, population growth)	WP4
	MAR yield predictions	WP4
	Aquifer characterization data	WP5
	Pollutant removal capacities from WW/SW	WP5
	LCA, LCC and S-LCA of technologies	WP6
DOC	Synthesis of (bio)materials	WP2
	Stakeholders' and citizens' information	WP6
	Information collected during the dissemination and communication	WP7
	Information collected during the exploitation regarding the knowledge generated within the project	WP7
SDF	Risk indicators GIS maps	WP4
	Aquifer characterization data	WP5

1.3   Data reuse

IHE Delft with permission of DU (Dunea Duin & Water) may be able to publish other information related to the location of Demo Site 1. All publications will be open access. NOVA with permission of AdTA (Águas do Tejo Atlântico) might share previous information related to the characterization of Demo Site 3. CIIMAR will revise all the data regarding contaminants obtained in the last years in the Lima River estuary (Demo site 7) in the frame of projects where they have been working on. These data will be obtained from published articles and public databases and will be used as a baseline for the following monitoring of the estuary in the frame of MAR2PROTECT.

Funded by the European Union

1.4   Origin of the data

There are two main origins of the data generated and collected during the project:

1- Open-source data available from scientific databases, namely meteo data, RCPs, sea level, land use & cover, and population growth.
2- New data generated during project implementation.

Furthermore, DU may generate other data that could be of relevance to the project. With their permission, IHE Delft may be able to use that data.

Internal and external data sources will be described in the metadata associated with research results.

1.5   The expected size of the data

MAR2PROTECT estimates a total data size of 1 TB for the whole project research data.

1.6   Data utility

The produced data will be useful to the MAR2PROTECT consortium partners and for dissemination, communication, and exploitation activities.

MAR2PROTECT research data will be useful to research institutions, water management bodies, legislators, etc, including, the general public interested in the improvement of water quality and quantity.

2 FAIR data

2.1 Making data findable, including provisions for metadata

MAR2PROTECT research data and metadata will be findable by:

- Referencing in the publications the unique identifiers for researchers by including the ORCID or other unique identifiers for researchers actually in use.
- Referencing digital object identifiers (DOIs) provided by data repositories when uploading a research result, in the research publications.
- Linking research publications DOIs with datasets DOIs.

A compilation of the MAR2PROTECT DOIs will be included in the **Published datasets** of the **DMP Data Annex**.

MAR2PROTECT metadata will follow the recommendations of the Metadata Standards Catalog[1] and will select the best metadata standards for the research that will be developed. Those standards, if any, will be included in the **DMP Data Annex.**

A compilation of the adopted or created metadata for the MAR2PROTECT project will be included in the **DMP Data Annex** and will be continuously updated.

In any case, metadata will cover all the next points:

- **Title**: Name of the dataset or research project that generated it. If the dataset is part of a manuscript, it is recommended (and often required) to use the manuscript title to link the data with the publication.
- **Creator/Author**: Names and addresses of the organizations or people who created the data. The preferred format for personal names is surname first (e.g., Ana B Pereiro). ORCID[2] IDs for authors will be used.
- **Funder**: Funding agencies/organizations, including the Crossref Funder ID[3].
- **Date**: Key dates, including project start and end dates, release date, the period covered by the data and other dates related to the data lifespan, such as maintenance cycle and update schedule. The preferred format is the ISO 8601 standard (e.g., yyyy-mm-dd or yyyy.mm.dd-yyyy.mm.dd for a range).
- **Description**: Search keywords or phrases describing the subject or content of the data.

---

[1] https://rdamsc.bath.ac.uk/
[2] https://orcid.org/
[3] https://www.crossref.org/

**Funded by
the European Union**

11

- **Place**: Note the physical locations where data are collected (e.g., NOVA University of Lisbon, NOVA School of Science and Technology FCT NOVA, Department of Chemistry, LAQV, REQUIMTE).
- **Method**: Describe how the data were generated, listing equipment and software used (including model and/or version numbers), formulae, algorithms, experimental protocols, reagents, and other details that one might include in a lab notebook. RRIDs[4] (Research Resource Identifiers) are recommended for citing key resources such as antibodies, model organisms, cell lines, plasmids, and other tools (e.g., software, databases, services…).
- **Processing**: A description of how the data have been filtered and processed prior to analysis.
- **Source**: Citations to data derived from other sources, including details of where the source data is held and how it was accessed.
- **File inventory**: A list of all files associated with the project, including extensions (e.g., baseline_CDI.csv, readme.txt)
- **File structure**: Directory URL where your datasets are located, along with a description of how data files are organized.
- **Necessary software**: List special-purpose software required to create, view, analyse, or otherwise use the data.

MAR2PROTECT-generated files will be named following the next structure:

Name_Reference_Title _Date_Version.Extension

- ***Name***: Project name.
- ***Reference***: Deliverable number (Dx.x) or Milestone number (Mx.x). If the file is not a deliverable or milestone, the Work Package number and Task number (WP1_T1.2) will be incorporated.
- ***Title***: Descriptive file name.
- ***Date***: Document date in ISO 8601 format (YYYYMMDD).
- ***Version***: Version number in a 3-digit (vn.n.n) format where:
  - The first digit is the big document version indicating a highly relevant change.
  - The second digit is a review indicator indicating the presence of minor changes.
  - The third digit indicated that a minor error or mistake was fixed.
- ***Extension***: the corresponding document extension.

---

[4] https://www.rrids.org/

**Funded by
the European Union**

All file name parts will be joined by an underscore and will have the next format:

MAR2PROTECT_Dn_title-or-other-content-specifications_YYYYMMDD_vn.n.n.Extension

MAR2PROTECT_WPn_Tn.n_title-or-other-content-specifications_YYYYMMDD_vn.n.n.Extension

A list of search keywords will be provided by the MAR2PROTECT project regarding data and publications and will be documented in the **DMP Data Annex**. This keyword list will be used by all partners.

2.2 Making data openly accessible.

Research data and outputs will be deposited and described in public data repositories that guarantee long-term preservation and can attribute persistent unique identifiers to the deposited items. Each partner will resort to its own data repository, making sure that all comply with the relevant EU requirements and FAIR practices. The following table shows each repository and the partner that will use it:

Table 2 Types of repositories for the MAR2PROTECT project.

Repository	Partner
ZENODO	NOVA, UNIBO, FEUGA, CIIMAR, CETAQUA, IHE DELFT, IT, ISSBAT, KTU, SU
RUN	NOVA
AMS Acta	UNIBO

MAR2PROTECT partners have selected the ZENODO repository as the project's open data repository. Project partners will use an internal repository not opened to the public and supported in SharePoint. This SharePoint will be used as a collaborative working place.

Repository	Partner
MAR2PROTECT SharePoint	NOVA, UNIBO, FEUGA, CIIMAR, CETAQUA, IHE DELFT, IT, ISSBAT, KTU, SU

Data produced and/or used in the MAR2PROTECT project will be revised and approved by all consortium partners before making it openly available. The communication of scientific publications will be carried out to the Consortium and the Exploitation Committee with 30 days' notice to evaluate if any information/results disclosure could injure a patent prosecution of confidential information agreements.

**Funded by
the European Union**

Furthermore, the deposit of the scientific publications must be carried out in a repository, immediately upon publication by the editor, without any embargo period[5]. Then, the only options are to choose diamond/gold open-access journals or hybrid journals (but selecting the gold open-access option). If a hybrid journal is chosen, the publication fee is not an acceptable cost for Horizon Europe projects. Another alternative is to publish in Open Research Europe[6], a publication platform created by the European Union. The project management team can support all partners, if it is necessary, in the selection of suitable journals that comply with the Horizon Europe rules, but at the same time are high-impact journals.

All data presented orally or as posters in scientific conferences and used to produce communication material, will be in line with the Horizon Europe guidelines and made available in open access through the beneficiaries' repositories. For this type of data, it is suggested that 15 days' notice will be given by any partner wishing to use project-related data to that effect.

The below table summarizes the different options:

Table 3 Notice required for each type of publication for the MAR2PROTECT project.

Type of Publication	Notice Required
Scientific Papers	30 days
Conference Proceedings	15 days
Website news items	15 days
Social Media Posts	15 days[7]

The consortium may decide that some data, including pieces of information, materials, and knowledge, should not be made openly available before IP rights protection or access is clarified. The Exploitation Committee shall be responsible for the dissemination, communication, and exploitation activities of the MAR2PROTECT project.

Research data underlying public reports and scientific publications will be deposited and made openly available immediately at the time of publication of results. The other project data will be deposited by the end of the project. All possible measures will be taken to ensure open

---

[5]   https://www.openaire.eu/how-to-comply-with-horizon-europe-mandate-for-publications

[6] https://open-research-europe.ec.europa.eu/

[7] Only in those cases where actual scientific data would be shared.

**Funded by
the European Union**

14

access to research data. Restrictions to the access of specific datasets will be applied only on account of privacy, ethical issues, copyright, confidentiality, and exploitation issues[8]. Data that cannot be openly shared will be specified in the DMP. The Data Protection Officer at NOVA (Dr João Marujo) will address any specific questions or issues regarding compliance with GDPR and related legal implications following the next decision tree.
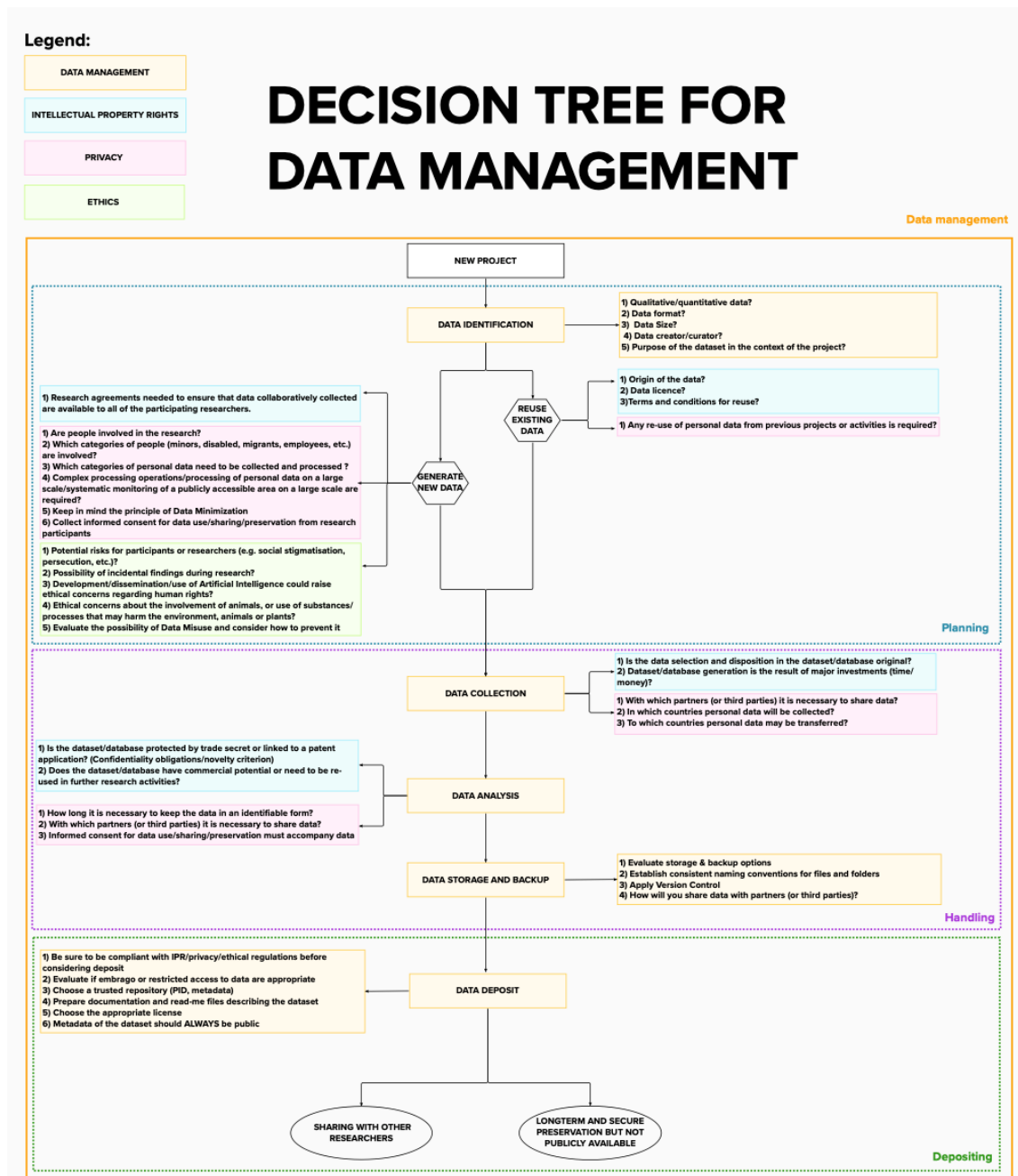

Figure 1 *Decision tree for data management (https://zenodo.org/record/7190005)*

All queries regarding scientific data storage and curation will be the responsibility of the General Assembly.

Publications resulting from project scientific achievements (public deliverables, communications to congresses and any other public document related to the MAR2PROTECT project) will be published as open access by default and deposited in the repository used by the relevant partner.

Data resulting from research work and related publications will be made accessible through the ZENODO repository and partners' specific repositories like RUN for NOVA and AMS Acta for UNIBO. These data will have an associated metadata description following previous recommendations.

ZENODO repository is maintained in CERN facilities with a 99.99% of availability.

ZENODO repository does not ask for user identification to access publicly available data, so no identification is requested. If the data are related to a scientific publication, while the publication is in its revision stage, access to data will be restricted and only allowed to reviewers.

The DPO and NOVA, as project coordinators, will assure that the Data Management Plan (DMP) is correctly implemented. Effective implementation will be guaranteed by the Project Management Team.

Only data with IPR implications will not be made public until finishing the IPR process and registered.

2.3 Making data interoperable

The concept of interoperable demands that data are correctly described in its metadata. Besides, it is stored in machine-readable files and documents where consistent terminology is used. Data produced in the MAR2PROTECT project will be made interoperable by a correct development of the metadata information and by using the most common data format possible. For example, for data files, the initial idea, while possible, is to use CSV-formatted ASCII files.

During the project, the use of open-source software is recommended, although, for certain tasks, licensed software can be used when needed. The final data will always be available in an open format, if possible because licenced software can use binary proprietary file structures. The associated metadata will include a description of the information and the application and version needed to open it.

Each partner will ensure that this procedure is compatible and in line with the requirements of their individual institutions' preferred repository.

**Funded by
the European Union**

MAR2PROTECT metadata will follow the recommendations of the Metadata Standards Catalog[9] and will select the best metadata standards for the research that will be developed. Those standards, if any, will be included in the **Adopted Metadata conventions**. of the **DMP Data Annex**.

Open data deposition across individual partners' data repositories will allow for inter-disciplinary interoperability in terms of metadata and vocabulary. For chemical compound terminology, the IUPAC[10] regulations will be followed.

Data and research outputs will be described using standard descriptive metadata as indicated in **Making data findable, including provisions for metadata**. The selected data repositories provide public APIs supporting OAI-PMH interoperability standards to facilitate automatic data harvesting.

Metadata will follow recommendations of the GO-FAIR[11] organization to improve interoperability.

2.4 Increase data re-use (through clarifying licences)

Open data will be made available according to Creative Commons Open Licenses that will be chosen when uploading data to the ZENODO repository.

The data will be available for re-use upon the decision of the Exploitation Committee, to avoid issues related to IP rights protection or access. Once the data are made openly available, it will remain open.

Within the strategy of DMP development, the datasets, that will be firstly available, are those reported in publications, originated from the consortium, and thus intrinsically made for being reused.

Each partner bears data responsibility and assures data quality. The tools necessary for describing and identifying the datasets and for preparing the metafiles will be provided by the General Assembly in strict collaboration with the Coordinator, the Exploitation Committee and the Data Management Officer.

3 Other research outputs

3.1 Allocation of resources

The project partners have allocated resources to cover costs associated with open-access publications. Partners will also use Open Access Publishing Platforms, e.g., the Open Research Europe platform.

---

[9] https://rdamsc.bath.ac.uk/
[10] https://iupac.org/
[11] https://www.go-fair.org/

**Funded by
the European Union**

The costs of data collection, quality check, cleaning and conversion to open formats, anonymization, pseudo-anonymization, description, and documentation can be estimated as 3.5% of the research activities costs. DMP-related activities will cost about 10 PM.

The CETAQUA cloud solution for secure storage during research entails an annual fee of 6 k€.

The costs for data storage during and after the project finishes are the responsibility of the Coordinator partner (NOVA) for internal data, and each consortium partner for open data. Open data will be deposited at ZENODO assuring its accessibility as it is archived permanently or at least during the CERN lifetime expected for the next 20 years.

4 Data security

MAR2PROTECT data will be stored in the ZENODO infrastructure. ZENODO is hosted by CERN which has existed since 1954 and currently has an experimental programme defined for the next 20 years. CERN is a memory institution for High Energy Physics and is renowned for its pioneering work in Open Access. Organisationally ZENODO is embedded in the IT Department, Collaboration Devices and Applications Group, Digital Repositories Section (IT-CDA-DR).

ZENODO is offered by CERN as part of its mission to make available the results of its work as stated in CERN Convention[12] , Article II, 1st paragraph.

About provided security:

CERN Data Centre: Our data centres are located on CERN premises and all physical access is restricted to a limited number of staff with appropriate training and who have been granted access in line with their professional duties (e.g., Zendo staff do not have physical access to the CERN Data Centre).

Servers: Our servers are managed according to the CERN Security Baseline for Servers, meaning e.g., remote access to our servers is restricted to Zenodo staff with appropriate training, and the operating system and installed applications are kept updated with the latest security patches via our automatic configuration management system Puppet.

Network: CERN Security Team runs both host and network-based intrusion detection systems and monitors the traffic flow, pattern and

---

[12]　https://council.web.cern.ch/en/content/convention-establishment-european-organization-nuclear-research#2

**Funded by
the European Union**

contents into and out of CERN networks in order to detect attacks. All access to zenodo.org happens over HTTPS, except for static documentation pages which are hosted on GitHub Pages.

Data: Zenodo stores user passwords using strong cryptographic password hashing algorithms (currently PBKDF2+SHA512). Users' access tokens to GitHub and ORCID are stored encrypted and can only be decrypted with the application's secret key.

Application: We are employing a suite of techniques to protect your session from being stolen by an attacker when you are logged in and run vulnerability scans against the application.

Staff: CERN staff with access to user data operate under CERN Operational Circular no. 5, meaning among other things that:

- staff should not exchange among themselves information acquired unless it is expressly required for the execution of their duties.
- access to user data must always be consistent with professional duties and only permitted for the resolution of problems, detection of security issues, monitoring of resources and similar.
- staff are liable for damage resulting from any infringement and can have access withdrawn and/or be subject to disciplinary or legal proceedings depending on the seriousness of the infringement.

ZENODO allows users to upload files under closed access. Closed access means that zenodo.org users will not be able to access the files you uploaded. The files are however stored unencrypted and may be viewed by ZENODO operational staff under specific conditions. This means that "closed access" on ZENODO is not suitable for secret or confidential data.

5 Ethical aspects

Ethical aspects do not apply to the MAR2PROTECT project.

6 DMP Data Life Cycle Annex

This Annex tries to condense all processes associated with the research data life cycle, including a checklist to be considered when making public data and its research results in a scientific publication.

6.1 Data Flow
The global flow of data and its relationship with scientific publications are shown in the next figure:
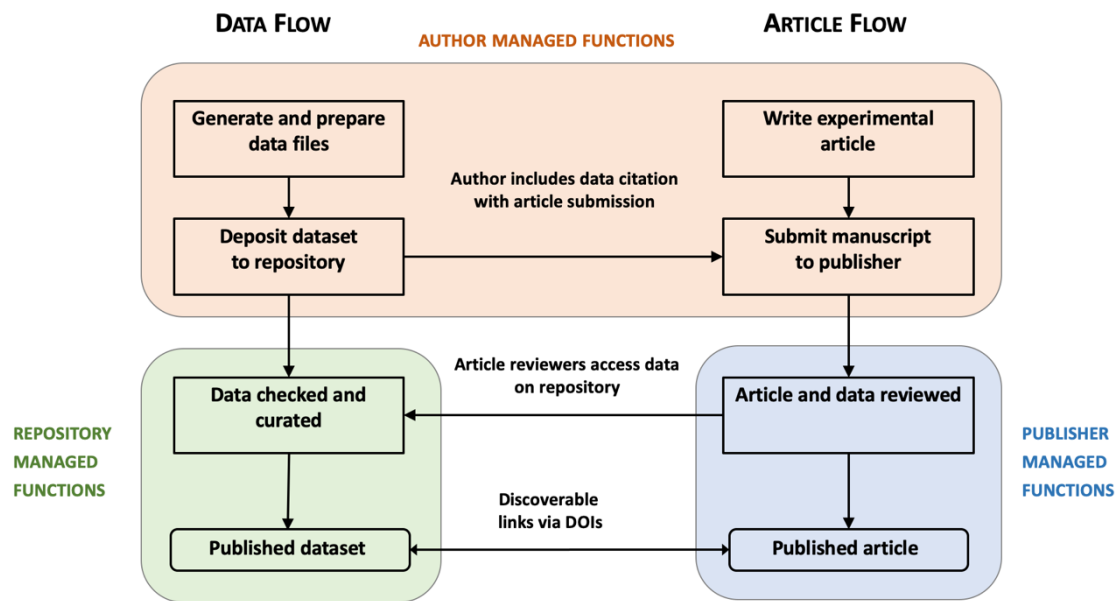
**Funded by
the European Union**

Figure 2 *The global flow of data and its relationship with scientific publications.*

Where each of the steps will be explained in detail in the next sections:

6.1.1   Generate and Prepare Data Files

- *Data Collection –* The first step is data collection/creation. Data needs to be collected from its origin and kept in a workspace (it is always recommended to make a backup).
- *Data Cleaning –* Data must be reviewed to be statistically correct, eliminating noise or identifying it and the same can happen with outliers, measurement errors, etc… This procedure ensures the quality of the data. After data cleaning, raw data, ready to work with, will be obtained for further storage.
- *Data Processing –* At this stage, the data can be analysed and processed. The analysis of the research data may also require the collection of new data for the same or other project purposes.
- *Data Storage –* Result data after the processing stage will enrich the initial raw data and generate final processed data that will be used to obtain research conclusions. These data will be stored for further processing, review or sharing purposes. In this stage, metadata will be generated by explaining the meaning of the collected data and the data obtained after processing raw data. This metadata will be stored in a separate document and included in the **DMP Data Appendix.**
- *Data Share –* Raw data and processed data (if different) will be uploaded to the project online repository and will be available to be accessed and discovered by third parties to be reused.

**Funded by
the European Union**

6.1.2   Deposit Dataset to Repository

In this step, data will be deposited in the ZENODO repository to be accessed by publication reviewers. The initial publication will be made restricted. Then, reviewers should ask for access to the data to check it.

Once data is uploaded to the ZENODO platform, a DOI is generated to be included in the paper before sending it to the publisher. The uploaded files cannot be modified anymore maintaining the same DOI, although new files like metadata files can be added. Versioning data will allow to update of the already uploaded data files.

ZENODO provides a special DOI that allows to reference all document versions called "Concept DOI". Then, if a document is possible to have multiple versions, like a software or an evolving dataset, refer to the Concept DOI for a general reference. This DOI is accessible after document publication and will reference the last published data (in review by Zenodo).

6.1.3   Write Experimental Article

Nothing to say to this point. MAR2PROTECT partnership has a highly qualified scientific background with long experience in publishing research papers. As data is uploaded to the repository, a DOI can be included in the paper to reference it.

6.1.4   Submit Manuscript to Publisher

No comment is needed.

6.1.5   Data Checked and Curated

After the paper review and data revision, reviewers will be able to access and check the research data. If a paper amendment is needed (e.g., needs more or updated data), a new version should be made.

6.1.6   Article and Data Reviewed

In this step, the publisher will notify the acceptance of the submitted paper for its publication.

6.1.7   Published Datasets

Once the article is accepted for publication, repository uploaded data can be made public before paper publication.

6.1.8   Published Article

Once the paper has been published, a final link can be included in the data repository to finally link data with publication.

6.2   Data FAIR CheckList

6.2.1   Datasets/files

● Is the dataset in the ZENODO repository?

● Does the dataset have a registered DOI?

**Funded by
the European Union**

● Are data files in standard and/or commonly available open formats (as much as possible)?

● Are the data and/or metadata retrievable via an API and/or discoverable through an open search protocol?

6.2.2 README/metadata

● Are all associated data files unambiguously named in the metadata and described including file types, software requirements and/or conversion information?

● Does the metadata include useful disciplinary notation and terminology? (e.g., SI units, common domain identifiers, explain acronyms, define field-specific jargon).

● Does the metadata include machine-readable standards where available (e.g. ORCIDs for authors and/or data contributors, W3C/ISO 861 date standard, and ITIS taxonomic IDs).

● Are related articles referenced and linked in the metadata?

● Is a citation format for the dataset provided?

● Are any license terms, attribution, or terms of use clearly indicated?

● Is the metadata exportable in a machine-readable structured text-based format? (e.g., CSV, XML, JSON).

6.2.3 Additional tips for preparing your data for sharing

6.2.3.1 *Preparing your data files:*

● You may choose to include raw data (as originally collected), processed data (e.g., signals encoded), or both. The decision depends on what is most useful or common in a discipline or specifically required by a publisher or repository.

● Use file formats that are common and open as much as possible, including for discipline-specific data types if open formats are available.

● Use unambiguous filenames and organize the files logically according to your project (e.g., by sample, treatment, method, etc.).

6.2.3.2 *Documenting your data and files:*

● For an easy, low-barrier approach, use a ReadMe template and save it as a plain text document (.txt).

● List the data files included in the package, and/or describe the file naming schema and organization. Include their formats and any specific software requirements and/or conversion information if you have it.

**Funded by
the European Union**

● Describe methods of data collection, file structures and organization including useful notation about the data headers, units, sample identifiers, etc. Use standard or conventional terminology or nomenclature in your discipline.

● Reference associated articles, code, and related datasets. Include ORCIDs of all data contributors.

6.2.3.3 *Depositing your data in a repository:*

● Make sure the repository provides a persistent identifier (e.g., DOI, handle, or other) and specifies conditions for others to access and re-use the data (such as a public domain declaration or Creative Commons attribution license; licensing policy may vary by the repository).

● Provide a pre-formatted citation (and license attribution if appropriate) for the dataset on your website and other materials so users can easily copy and attribute, for example:

- Author(s). Dataset Title, Version. Data Repository (or Journal if appropriate). Year. DOI. (Date accessed).
- [License attribution if appropriate].

7 DMP Data Annex

7.1 Adopted Metadata conventions.

This is the list of adopted metadata conventions:

Convention	Description

7.2 Published datasets

Data Management Plan document will be updated with the list of published datasets mostly corresponding with ZENODO published datasets but with the possibility to include datasets published by partner institutions internal repositories (non-mandatory).

For an easier management, a table in the MAR2PROTECT SharePoint will be added with the same fields.

Dataset Description	Type of PID[13]	PID	URL	Status[14]	Partner	Pub Date

---

[13] **Type of PID**: Will indicate the repository origin of the Persistent ID: ZENODO, RUN, AMS Acta or any other.

[14] **Status**: Will indicate the repository status: Open, Embargoed, Restricted, Closed.

7.3   Keywords of published data

Document name	DOI	Keywords

**Funded by
the European Union**