DELIVERABLE D6.6 - DATA MANAGEMENT PLAN

(To be filled in and uploaded as deliverable in the Portal Grant Management System, at the due date foreseen in the system (and regularly updated).

⚠ *The template is recommended but not mandatory. If you do not use it, please make however sure that you comply with the research data management requirements under Article 17 of the Grant Agreement.)*

PROJECT	
Project number:	101079345
Project acronym:	SURRI
Project name:	Sustainable Remediation of Radionuclide Impacts on Land and Critical Materials Recovery

DATA MANAGEMENT PLAN	
Date:	16.12.2025
Version:	1.2

1. Data Summary

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

The project primarily generates new data and also uses site-specific data provided by DIAMO a.s. under contractual arrangements. These data are necessary to meet the project objectives and include partially sensitive information related to the project sites. There is no indication that additional external or open datasets are re-used beyond the data generated within the project and the data obtained from DIAMO.

What types and formats of data will the project generate or re-use?

The project generates several types of instrument and experimental data, including real-time PCR, next-generation sequencing, ICP-MS, ICP-OES, gamma spectrometry, and field measurements of physical-chemical water parameters such as temperature, pH, conductivity, and ORP. The data are stored in standardized formats suitable for analysis and long-term use, including FASTQ, ixo, xlsx, CSV, and docx. Additional formats include tabular, text, image, and structured data formats (e.g. xml and json).

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

The purpose of data generation is to measure, analyze, and process samples collected in the field and laboratory using the measuring instruments of TUL and other project beneficiaries, as well as data from sites provided by DIAMO. The data support scientific analysis, validation of results, development of technologies, and preparation of scientific publications, thereby directly contributing to the objectives of the project.

What is the expected size of the data that you intend to generate or re-use?

The expected size of the generated and re-used data is up to maximum 100 GB in total.

What is the origin/provenance of the data, either generated or re-used?

The origin and provenance of the data are clearly defined. Data are generated by project experts using their own well-described and well-known equipment and collected through field measurements and laboratory experiments. Data provenance is ensured through the use of laboratory notebooks, standard operating procedures for file naming (including site name, sampling date, and data type), and metadata describing instruments, settings, and sampling conditions. Data obtained from DIAMO are governed by contractual agreements that define access and sharing conditions.

To whom might your data be useful ('data utility'), outside your project?

Outside the project, parts of the generated data – particularly real-time PCR, next-generation sequencing, and gamma spectrometry datasets – may be useful to other researchers working in the same field. Published data may also be relevant for other research institutions or companies for further scientific analysis, statistical evaluation, and optimization of practices and technologies. Access to the data is subject to legal and contractual constraints, especially for site-specific data provided by DIAMO.

2. FAIR data

2.1. Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

The data generated within the project will be identified using persistent identifiers, in particular DOIs (Digital Object Identifiers), which are assigned when data are deposited in the Zenodo. This ensures long-term findability and traceability of the datasets.

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Rich metadata are provided to allow data discovery and interpretation. Metadata include information on sample descriptions, analytical results, laboratory experiments, instrument settings, software required to read raw data, and relationships to publications and other project outputs. Although no formal minimal metadata requirements exist for the specific experiments carried out in the project, sufficient metadata are created to enable understanding and reuse of the data by other researchers.

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Samples are labeled according to the place of collection, date of collection, and type of sample. It will then be possible to search this data using these keywords, which also link to the experiments.

Will metadata be offered in such a way that it can be harvested and indexed?

Metadata are offered in a way that allows them to be harvested and indexed by repositories. The Zenodo Repository, compliance with the DataCite metadata standard, and assignment of persistent identifiers enable indexing, linking to publications, and long-term discoverability, even in cases where access to the underlying data is restricted or data are no longer available.

2.2. Making data accessible

Repository:

Will the data be deposited in a trusted repository?

The data generated within the project will be deposited in trusted repositories, primarily the Zenodo repository operated by CERN and supported by the European Commission and storage services provided by the CESNET e-infrastructure. During the project, data are shared and backed up using an online drive (Google Drive) with controlled access for project members. Upon completion of the project, long-term storage is ensured using CESNET S3 storage or the Zenodo repository.

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

Appropriate arrangements with the identified repositories have been explored. The project follows existing institutional and national infrastructures, which support long-term archiving, persistent identifiers, and compliance with FAIR principles.

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

The repositories used ensure that datasets are assigned a persistent identifier (DOI). These identifiers resolve to digital objects and enable long-term findability, traceability, and linking to related publications and other project outputs.

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

Not all data generated within the project will be made openly available. Restrictions apply primarily due to legal and contractual reasons, particularly agreements with DIAMO a.s., which provides access to the project sites and supplies partially sensitive site-related data. Under these contractual arrangements, DIAMO has the right to restrict or prevent public release of specific datasets.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

In addition, there are intentional restrictions related to the need to publish scientific results and, where relevant, to protect institutional know-how. Full disclosure of data generally takes place after publication or patenting, where applicable. Some restricted data are embargoed indefinitely, and therefore limited embargo periods cannot always be applied.

Will the data be accessible through a free and standardized access protocol?

When data are made available, they are accessible through free and standardized access protocols, typically via repository URLs that allow users to download the data without fees. Once data are open, there are no restrictions on their use, and no additional identifiers or authorization are required.

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

During the project, access to data is restricted to project investigators through controlled storage environments. After the project, access to data depends on contractual and legal conditions. Open data are freely accessible, while restricted data remain accessible only to authorized parties under the terms defined in the consortium and contractual agreements.

How will the identity of the person accessing the data be ascertained?

The identity of users accessing openly available data is not required to be ascertained, as no authorization is needed. For restricted data, access is managed contractually, and there is no need for a data access committee, as the data do not contain personal data and access conditions are clearly defined in existing agreements.

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

The sharing of individual data is contractually arranged with DIAMO, which provides the sites needed for the project. DIAMO has access to the data and reserves the right not to share the data publicly.

Once the data is open, there will be no restrictions on the data. The data will be freely available. The repository will have an url for the user to download the data. There will be no need for a data access committee.

Metadata:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

Metadata are made openly available and are licensed under CC0. The metadata contain sufficient information to enable users to understand how the data can be accessed, including repository location, identifiers, and access conditions.

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

The data remain available and traceable for five years after the end of the project. Metadata remain available even after the underlying data are no longer accessible, ensuring continued findability and contextual understanding of the datasets.

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

Documentation describing any special software required to access or read the raw data is included in the metadata. This includes information on proprietary software and its version used to generate or process the data. Because the required software is proprietary, the project provides processed and exported data alongside the raw data.

The inclusion of the software itself (e.g. as open-source code) is not described in the document and therefore not foreseen.

2.3. Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

The project ensures data interoperability by using standardized and widely accepted data formats. The main formats used include FASTQ for sequencing data, xlsx and CSV for tabular data, docx for documentation, and structured formats such as xml and json. These formats are commonly used across disciplines and support data exchange and reuse.

For metadata, the project follows the **DataCite Metadata Schema v4**, which is a widely adopted general standard supporting interoperability, persistent identifiers, and linking between datasets and other research outputs. The use of this schema aligns with community-endorsed best practices for data interoperability and integration within trusted repositories.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

No specific ontologies or vocabularies should be created within the project.

Will your data include qualified references[1] to other data (e.g. other data from your project, or datasets from previous research)?

The data include qualified references to other data and outputs through metadata and repository practices. Metadata link datasets to related publications, project identifiers, authors, and other project outputs, enabling users to understand relationships between datasets produced within the project. If other data from the project or from previous research are used, they will contain references where applicable.

[1] *A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/)*

2.4. Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

Documentation needed to validate data analysis and facilitate data re-use is provided through rich metadata, laboratory notebooks, and repository documentation. Metadata include detailed information on sample descriptions, analytical methods, instrument settings, software required to read raw data, and relationships to other datasets and publications. Standard operating procedures (SOPs) for file naming ensure clarity regarding site, sampling date, and data type, supporting data interpretation and reuse.

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

The data generated in the project cannot be made fully and immediately available in the public domain. Restrictions apply mainly due to legal and contractual obligations, especially agreements with DIAMO a.s., as well as the need to publish scientific results and, where relevant, protect institutional know-how. Where possible and appropriate, data are made openly available after publication. For open datasets, the project assumes the use of standard reuse licenses, in particular CC-BY 4.0 or CC0, in line with the Grant Agreement.

Will the data produced in the project be useable by third parties, in particular after the end of the project?

Once data are openly released, they are freely available and reusable by third parties, including after the end of the project, without restrictions or the need for individual authorization. Even when data cannot be fully disclosed, metadata remain available to support discovery and potential controlled reuse.

Will the provenance of the data be thoroughly documented using the appropriate standards?

The provenance of the data is thoroughly documented using laboratory notebooks, SOPs, metadata standards, and repository practices. Metadata capture information on data origin, instruments used, analytical conditions, software versions, and processing steps, ensuring transparency and traceability.

Describe all relevant data quality assurance processes.

Data quality assurance is addressed through several processes. Data are generated by experienced project experts using well-described and known equipment. To validate the integrity of results, parts of the datasets are run repeatedly to detect unexpected changes in outcomes. Data are archived during and after the project using long-term storage (“cold storage”), supporting data integrity and long-term reuse.

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

Beyond FAIR principles, the project also addresses other research outputs, allocation of resources, data security, and ethical aspects. Data stewardship is ensured through a designated Data Steward role, supported by trained staff. Secure storage infrastructures are used, risks related to data loss or leakage are considered low, and no personal data are processed. Ethical aspects are addressed through project-level ethical review and compliance with applicable legislation.

2.5. Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

We use Data Stewardship Wizard for planning our data management and creating this DMP. Still, we benefit from data stewardship guidance (e.g. FAIR principles, openness, or security) and it is reflected in our plans with respect to other research outputs.

In addition to research data, the project generates physical research outputs, namely samples of mining sludges and mine waters collected at project sites. These samples are used as input materials for laboratory and experimental activities focused on remediation and recovery processes.

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

The physical samples are collected, stored, and handled according to established laboratory and institutional procedures. They are not intended for public distribution or long-term external reuse, as their purpose is to support experimental work within the project and to enable the generation and validation of research data.

Information describing these physical research outputs, including sampling location, date, sample type, and subsequent experimental use, is documented through metadata, laboratory notebooks, and internal documentation. This ensures transparency, traceability, and reproducibility of the research results derived from these samples.

Where applicable, FAIR principles are applied to the documentation describing the samples, rather than to the physical samples themselves. Access to the samples and related documentation is subject to legal and contractual constraints, particularly agreements with the site owner (DIAMO a.s.). The results obtained from experiments using these samples are shared through scientific publications and associated datasets, in line with the project's data management and dissemination strategy.

2.6. Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?

FAIR data management is an integral part of the project and is considered at every decision related to data handling. Making data FAIR is not treated as a separate cost but as part of the overall project activities.

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

The project uses repositories and infrastructures that do not charge fees for storage or archiving, including the Zenodo repository and CESNET infrastructure. Cloud storage is provided free of charge through CESNET membership.

Who will be responsible for data management in your project?

Martin Palusak is responsible for maintaining the finished data resources.

Jakub Zach is responsible for data management and proficiency, including data processing, data policies, data guidelines, and data availability.

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

Long-term preservation is ensured through long-term (“cold”) storage, both during and after the project. Data are archived using trusted infrastructures, and decisions on preservation are made within the project according to legal, contractual, and scientific value considerations. Data remain available for at least five years after the end of the project, while metadata are retained for longer.

3. Data security

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

Data security is ensured through a combination of secure storage, controlled access, and certified infrastructures. Data archives are stored in remote locations to protect against disasters, loss, or theft. It is clearly defined who has physical and logical access to the data.

Will the data be safely stored in trusted repositories for long term preservation and curation?

Project members do not store data or software on laboratory computers, external hard drives, laptops, or portable media such as USB sticks. All project web services use secure communication protocols (HTTPS), and data centers used for storage carry sufficient certifications.

The potential impact of data loss, leakage, or vandalism is considered low. No personal data are processed in the project. Trusted repositories and infrastructures are used to ensure long-term preservation and secure archiving of the data.

4. Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

The data produced and collected within the project do not include personal data. Collected sample descriptions and analytical results do not contain sensitive data, while laboratory experiments may contain sensitive, non-personal data. All data collection activities comply with applicable ethical legislation and are covered by an ethical review.

Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

The project does not involve human subjects, and therefore informed consent for data sharing and long-term preservation is not applicable. Ethical and legal considerations affecting data sharing are mainly related to contractual obligations with DIAMO and the handling of sensitive site-related information.

5. Other issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

The project does not use additional national, funder-specific, sectorial, or departmental data management policies or procedures beyond those already described. Data management follows established project, institutional, and repository practices, in line with the Horizon Europe Grant Agreement and FAIR principles.

HISTORY OF CHANGES		
VERSION	PUBLICATION DATE	CHANGE
1.0	05.05.2021	Initial version (new MFF).
1.1	01.04.2022	Reformatted to align with other deliverables templates.