MADITRACE

Data Management Plan

Deliverable D6.5

Version N°4

Authors: Daniel Monfort (BRGM), Mariana Terreros (LGI), Samy Strola (LGI)

Contributors: A. Desaulty (BRGM), M. Katapotis (DMT), L. Forster (DMT), P. Jussi (GTK), M. Kugerl (MUL), F. Vanhaecke (Ugent), I. Calleja (EIT RM), M. Fernandez (ISMC), A. Garcia (ISMC), M. Montes (ICAMCYL), Agnès Lanusse (CEA), L. Donnelly (AHK), A. Garcia (Funditec), S. Cunha (ULEI), J. Kaartinen (MO Group), J. Ebert (Spherity), Adeline Paul (LGI)

Funded by the European Union

Disclaimer

The content of this report reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

Document information

Grant Agreement	101091502
Project Title	Material and digital traceability for the certification of critical raw materials
Project Acronym	MaDiTraCe
Project Coordinator	Daniel Monfort, BRGM
Project Duration	1 January 2023 – 31 December 2025 (36 months)
Related Work Package	WP6 Project and data Management
Related Task(s)	T6.5
Lead Organisation	LGI
Contributing Partner(s)	BRGM, DMT, GTK, MUL, UGent, EIT RM, LGI, ISMC, ICAMCYL, CEA, AHK, Funditec, ULEI, MO Group, Spherity.
Due Date	M3
Submission Date	27/06/23, resubmitted 10/02/25
Dissemination level	Public

History

Date	Version	Submitted by	Reviewed by	Comments
23/03/23	V0.1	Daniel Monfort	Mariana Terreros	
02/05/23	V0.2	Mariana Terreros	Samy Strola	
03/05/23	V0.3	Mariana Terreros	All partners	
15/05/23	V0.4	Mariana Terreros	Daniel Monfort	
27/06/23	FINAL	Mariana Terreros	Daniel Monfort Samy Strola	Submission 1
26/11/24	V1.0	Adeline Paul	Daniel Monfort	
03/12/24	V1.1	Adeline Paul	All	
10/02/25	V1.2	Adeline Paul	EC	Submission 2

Table of contents

List of tables

Summary

The MaDiTraCe Data Management Plan (DMP) is based on the Horizon Europe (HE) DMP template that was designed to be applied to any HE project that produces, collects or processes research data. It presents the status of the data that is collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved.

This DMP lists the data management principles, strategies, tools and types of data that will be produced as part of the project activities. The consortium will also aim at open access when publishing papers and articles.

This deliverable provides an overview of research data the MaDiTraCe project is expected to generate, the types and formats of this data, and how this data is processed and stored to follow the principles of FAIR data management (findable, accessible, interoperable and reusable). The purpose of the DMP is to contribute to good data handling during the project lifetime.

The DMP is a living document that will be updated as the implementation of the project progresses and when significant changes occur. The current version corresponds to the second submission, produced 25 months after project start and after the first periodic report. The report is now more in line both with MaDiTraCe current issues and with the reference DMP template.

Keywords

Data management, findable, accessible, interoperable and reusable.

Abbreviations and acronyms

Acronym	Description
CA	Consortium Agreement
CC	Common Creative
CERA 4in1	Certification of Raw Materials
CPS	Cyber Physical Systems
CSS	Customer Service and Support
D	Deliverable
DMP	Data Management Plan
DPO	Data Protection Officer
DoA	Description of Action
DOI	Digital Object Identifier

EC	European Commission
FAIR	Findable, Accessible, Interoperable and Reusable
GA	Grant Agreement
GDPR	General Data Protection Regulation
HE	Horizon Europe
IPR	Intellectual Property Rights
ISSN	International Standard Serial Number
LCA	Life Cycle Analysis
M	Month (of the project)
NDA	Non- Disclosure Agreement
PII	Publisher Item Identifier
PMO	Project Management Office
PU	Public
ORDP	Open Research Data Pilot
SEN	Sensitive
USGS	Unoted States Geological Survey
URL	Uniform Resource Locator
V	Version
WP	Work Package
WPL	Work Package Leader
Gold Open Access	Open access publishing (gold open access) means that an article is immediately provided in open access mode on the publisher or journal's website. Some publishers charge Article Processing Chargers (APCs) to make articles open
Green Open Access	Self-archiving (green open access) means that a published article or the final peer-reviewed manuscript is archived (deposited) in an online repository before, alongside or after its publication. In some cases, the author can choose to delay access to the article (embargo period). HE rules state that embargo periods cannot exceed six months, except for publications in social science and humanities where the maximum embargo period is twelve months
Personal Data	Personal data is any information that relates to an identified or identifiable living individual. Different pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data. Personal data that has been de-identified, encrypted or pseudonymised but can be used to re-identify a person remains personal data and falls within the scope of the law. Personal data that has been rendered anonymous in such a way that the individual is not or no longer

	identifiable is no longer considered personal data. For data to be truly anonymised, the anonymisation must be irreversible
Zenodo	Zenodo is a catch-all research data repository that enables researchers, scientists, EU projects and institutions to share research results, make research results citable, and search and reuse open research results from other projects. Zenodo is harvested by the OpenAIRE portal and hosted by the CERN cloud infrastructure

1 Introduction

1.1 Intended readership / users

Internally in the project:
- All project participants who are responsible for, or in any way involved with, data collection and data handling shall refer to this document for instructions on how to handle, store and process data.
- All project participants can use this document to get an overview of all data collected in the project and how it is processed and stored.

External audience:
- All relevant stakeholders who are interested in MaDiTraCe activities and research topics can use this document to get an overview of the data collected in the project, how to access this data, and, if applicable, how to re-use this data in their own activities.
- All persons who voluntarily participate in the project (4 commodities cases studies, DPP, certifications) and contribute data to the project can use this document to learn how the project processes and stores their data.

1.2 Objectives and structure of the document

The DMP describes the data management life cycle for the data to be collected, processed and/or generated by MaDiTraCe. The DMP aims at defining the management strategy of data generated during the project with the purpose to making research data findable, accessible, interoperable and re-usable (FAIR). The document is structured following the guideline of HE programme on FAIR Data Management including the following information:
- DMP guiding principles and strategy
- Description of MaDiTraCe type of data
- Description of FAIR DATA characteristics including DMP Review Process & data inventory
- Data Security
- Ethical considerations
- Allocation of resources
- Conclusions

According to the EU's guidelines regarding the DMP (European Commission, 2016), the document may be updated - if appropriate - during the project lifetime (in the form of deliverables). The minimum requirement is that the DMP is updated for each periodic evaluation of the project and in time for the final project review.

The DMP is intended to be a living document in which information can be further detailed as the implementation of the project progresses.

The report therefore has a clear version number and include a timetable for updates.

2 Data Summary

2.1 Re-use of existing data and purpose

Partner	Existing data that will be used or re-used
BRGM	Data from previous works (Li traceability paper of Desaulty et al 2022). Other existing data available in BRGM such as data generated by the Mineral Intelligence Unit.
DMT	Data from CERA-project, funded by EIT Raw Materials (2017-2021). Basis for standard and audit catalogue development.
GTK	Data from previous work will be reused, including the Business Finland Co-Research project BATTRACE, which involved creating a geochemical reference library for the purpose of material traceability, based on both published analyses and analyses conducted during the project.
MU Leoben	Existing data from previous projects (e.g. Analytical proof of origin for natural graphite deposits (PhD thesis Valentina Dietrich), CERA4in1 will be (re)-used. Raw and processed data of these projects might be (re)-used. Already published data will not be used in this project again.
UGent	Existing data from previous projects under the form of analytical background information.
EIT RM	Existing data available at EIT RM.
LGI	No data from previous work will be used in this project.
ISMC	No data from previous work will be used in this project.
ICAMCYL	No data from previous work will be used in this project.
CEA	No data from previous work will be used in this project.
AHK Group	Selected data will be reused from AHK's pre-MaDiTraCe project laboratory investigations that determined the physical and chemical characteristics of artificial taggant particles. This may include automated mineralogy, optical mineralogy, X-CT and LA-ICP-MS.
Funditec	No data from previous work will be used in this project.
ULEI	Domestic extraction data of minerals, as well as trade and processing, including necessary data to create a dynamic material flow accounting model for the secondary flows (paper by Susan van den Brink et al. https://doi.org/10.1016/j.resconrec.2020.104743, and master thesis by Tomás Joaquín Saiz Vinet) . Life-cycle inventory data, where available, e.g. for battery raw material supply chains (PhD theses Marc Meide and Carina Harpprecht).
MO Group	No data from previous work will be used in this project.
Spherity	Some of the existing data listed above if shared by other consortium partners and relevant for the data exchange system.

Table 1 List of data that will be used or re-used by partners

For more information see Annex I, 8.1 Data Collection.

2.2 Types and formats of data generated

Partner	Type	Format
BRGM	Both published and self-generated data on analytical measurements. Maps of samples analysed. Pictures/images. Reports.	.docx .xlsx, csv, tsv .pptx GIS: shp WMS, WFS, Atom JSON.LD files Images: jpg, png, tiff Pdf Cryptographic data, such as signatures, proof requests, etc.
DMT	Standard documents and audit catalogues / presentations.	xlsx, docx, pptx
GTK	Both published and self-generated data on analytical measurements. Pictures/images. Report.	xlsx, docx, jpg, tiff, pdf
MU Leoben	In the framework of literature review (T&T compliance) a report will be generated. In the framework of the performed analytics, a distinction must be made between different types of data. Raw data, processed data and secondary data outputs, such as reports will be generated.	.docx .xlsx, csv, tsv, h5 .pptx .txt .pdf .rep .xl Jpg, png, tiff cdr, io
UGent	Data will be generated via measurements with ICP-MS and MC-ICP-MS instrumentation. Data will be processed using instrument-specific software and Excel. Statistical software will be used to aid interpretation of data. Results will be summarized in text format, tables, graphs, PowerPoint presentations.	Instrument-specific file formats .xlsx .csv .tsv .docx .jpg .tiff .pdf .pptx
EIT RM	Stakeholders database.	.xlsx
LGI	Consortium contacts. Registration form for meetings/workshops through https://evenium.events/ (temporary use)	.xlsx .xlsx
ISMC	Report. Pictures/images.	.docx .xlsx, csv, tsv

	Charts or tables, List of subscribers to the project newsletter in Mailjet	.pptx .pdf
ICAMCYL	Pictures/images. Charts or tables.	.docx .xlsx, csv, tsv
CEA	Not applicable. No data collection. Documents exchanged for demonstrations will be dummy documents and/or public and generated data provided by the consortium partners.	.docx .xlsx, csv, tsv .pptx .pdf .xml
AHK Group	AHK generated and existing published data and information.	Xlsx, docx, jpg, tiff, pdf
Funditec	Not applicable.	
ULEI	Published and generated datasets. Relevant results will be included in a report and possible presentations.	.docx .xlsx, csv, tsv .pptx .pdf
MO Group	Raw data provided by MO Group's Courier® analyzer(s).	Instrument-specific file format
Spherity	Both public and generated data provided by the consortium partners.	.json .pdf JSON-LD Cryptographic data, such as signatures, proof requests, etc .csv

Table 2 Types and formats of data generated/collected by partners

For more information and guidelines see Annex I, 8.1 Data Collection and 8.2 Data Description.

2.3 Purpose of the data generation and its relation to the objectives of the project

Partner	Purpose data that will be collected/generated for the project
BRGM	WP2: Geochemical database of lithium will be used as a reference library when distinguishing between material originating from different lithium bearing deposits. Inter-lab comparisons with some of the project partners require production of various mineralogical and geochemical data by using different instruments. Algorithms for data treatment, statistics analysis and machine learning. WP1-3-4: The publicly available data (USGS, BGS, and company websites), commercial databases, and UN Comtrade will be used to make the supply chain mapping. The life cycle inventory data will be used to determine the LCA of primary and secondary value chains of key materials and products that will then serve to quantitatively describe the production process and how this could affect the traceability of the materials. The LCA

Funded by the European Union

	will also be used to calculate the environmental footprint of the material supply chains that will serve as inputs to the CPS and CCS.
DMT	The already existing CERA 4in1 Performance Standard and its corresponding documents will be used as guideline to develop the other supply chain standards within the MaDiTraCe project. As part of the project, pilot audits will be conducted at selected sites of companies to evaluate in practice the performance of the developed CERA standards. In this process, company data will be collected through inspections, interviews and company-document reviews, primarily relating to ESG procedures.
GTK	A geochemical database of cobalt will serve as a reference library for distinguishing material originating from different cobalt-bearing deposits. Inter-laboratory comparisons with some project partners will necessitate generating diverse mineralogical and geochemical datasets using various instruments.
MU Leoben	The sustainability compliance database will be used to identify gaps and assess the needs of the manufacturing industry. On the other hand, the MFP database will be used to distinguish between different economically relevant deposits in the first step and to track and trace the raw materials along their value chain in the second. Both studies will then be shared with third parties involved in the project and the industry.
UGent	Within WP2 UGent will develop analytical protocols for elemental fingerprinting and Nd isotopic analysis. These analytical protocols comprise sample preparation, measurement, data processing and chemometric data treatment and will be used for the characterization of "magnet-related" materials from various geographic locations and from various stages of the chain of custody. The results obtained in this context will be collected in a database. Inter-lab comparisons with some of the project partners require production of various mineralogical and geochemical data by using different instruments.
EIT RM	Stakeholders database
LGI	No tasks envisions the collection of data to support the project results but rather to support the monitoring and promotion of the project itself.
ISMC	The C&D data collection will be used to collect information on the C&D activities carried out in the project. This data collection tool will facilitate the planning and timing of news and publications that will be visible on the project's communication channels. Basically, it will allow us to analyse the data through KPIs measurements aiming to evaluate the impact of the project through the production of a report.
ICAMCYL	Clustering activities will collect data from related projects and assistants to relevant project events. This data will be used to facilitate contact with the relevant projects and clustering event assistants, and to keep a register of the public reached.
CEA	Not applicable. In WP3 CEA will not collet/generate Data but will provide demonstration based on "dummy examples" to demonstrate interest of various techniques proposed by WP3 (blockchains, DD, dedicated smart-contracts etc...). Possible application to real data will be done within use cases under the responsibility of use case provider. In WP2 CEA will provide infrastructure for experimentation on material samples. Data analysis will be achieved by BRGM.

	Publication on methodologies applied and their implementation on typical use case are envisaged.
AHK Group	Data and information will be collected on the applications of 3D printing technology and artificial taggant particles that have been used for the traceability of non-mineral/metal commodities, and then subjected to review. Most of the data and information will be from publicly available on-line sources or from suppliers of artificial taggants. These data will be analysed to determine their physical and chemical characteristics, and to determine whether these can be adopted or adapted for applications to the MaDiTraCe target minerals/metals.
Funditec	The purpose of the activities of FUNDITEC do not involve data collection but issuing recommendations about security, confidentiality, accountability at design and runtime control of the raw materials data collected by MaDiTraCe partners. However, FUNDITEC will issue a report to the related missing controls or issues found in such data collection/generation for the purposes of the project and in particular task T3.4 and deliverable D3.7.
ULEI	The publicly available data (USGS, BGS, and company websites), commercial databases, and UN Comtrade will be used to make the supply chain mapping. The life cycle inventory data will be used to determine the LCA of primary and secondary value chains of key materials and products that will then serve to quantitatively describe the production process and how this could affect the traceability of the materials. The LCA will also be used to calculate the environmental footprint of the material supply chains that will serve as inputs to the CPS and CCS.
MO Group	MO Group will receive mineral samples and generate measurement data (e.g. spectra containing characteristic peaks for various elements). The data will be generated with MO Group's proprietary Courier® analyzers and stored by MO Group.
Spherity	No self-generated data. The public and generated data from other work packages and tasks, as well as dummy data for product information, raw material description and organisational data will be used to design and create a data exchange platform and Digital Product passport.

Table 3 Purpose of data collection and generation for the project by partners

For more information and guidelines see Annex I, 8.1 Data Collection and 8.2 Data Description.

2.4 Expected size of the data we intend to generate or re-use

Partner	Size of data generated or re-used
BRGM	>10 GB. Mineral characterization and spectra can generate very big files.
DMT	Data sizes will be small mainly text based (standard documents and audit catalogues).
GTK	<10 GB

MU Leoben	Most of the generated data will be raw and processed data (bigger data sizes), however most of it will not be disseminated in general, but only compared with other laboratories. The transfer of the data, however is still easy to handle. Secondary data (e.g. reports) is mainly text based and showing statistics, easy to distribute and the estimated sizes are rather small. No additional costs need to be estimated here for data transfer and storage.
UGent	<10 GB
EIT RM	Not applicable.
LGI	Not applicable.
ISMC	Not applicable.
ICAMCYL	Not applicable.
CEA	Not applicable.
AHK Group	Relatively small file sizes that are KB and MB, but some could extended to GB.
Funditec	Not applicable.
ULEI	Small file sizes (KB / MB rather than GB).
MO Group	Raw data generated by the analyzer(s) is in the GB range.
Spherity	Data sizes will be small mainly text based.

Table 4 Size of data generated or -reused by partners

For more information and guidelines see Annex I, 8.3 Data volume.

2.5 Origin/provenance of the data, either generated or re-used

The MaDiTraCe project will generate data after chemical analyses of samples of different minerals, products and by-products coming from Cobalt, Lithium, Natural Graphite and Neodyme supply chain. These samples will come from already existing and available samples in each partner but also from new samples obtained during the project. For many of them the project will contact industry and mining companies to provide samples. The conditions of reuse of obtained data from these samples will depend on the discussion between the MaDiTraCe project and each simple provider. Each partner engaged in the obtention of samples has detailed the origin and provenance of data in the following table:

Partner	Origin of data that will be generated or re-used
BRGM	re-used data from previous R&D projects (EIT RM project called EuGeLi, BRGM-EDF project about lithium traceability and the published work of Desaulty et al. (2022)). Samples obtained by field campaigns of BRGM in previous projects. BRGM will ask permissions from the sample providers concerned if applicable.
DMT	DMT uses data and background expertise about certification in mining sector gained from the EIT Raw Materials funded project CERA – Certification of Raw Materials (2017-2021). New data will be generated based on desktop research and communication activities with other Initiatives that are developing certification schemes as well. More over sensitive data of pilot partner/companies will be gained during the pilot testing of developed standards.

GTK	1) Data from previous work will be reused, including the Business Finland Co-Research project BATTRACE, which involved creating a geochemical reference library for the purpose of material traceability, based on both published analyses and analyses conducted during the project. 2) MaDiTraCe consortium intends to reanalyze some of the BATTRACE samples, and GTK will request permissions from the respective sample providers. 3) GTK conducts analyses on new samples acquired during the MaDiTrace project.
MU Leoben	The generated data (new and re-used) come from in-house research facilities and laboratories, or were generated by other project partners. If analyses are outsourced, the origin of the data is identified.
UGent	The data will be generated using measurement of elemental concentrations using ICP-MS and isotope ratios using multi-collector ICP-MS. Raw data will be processed and for interpretation, chemometric approaches will be relied on.
EIT RM	Existing data available at EIT RM.
LGI	Not applicable.
ISMC	Not applicable.
ICAMCYL	Not applicable.
CEA	Not applicable.
AHK Group	Existing data and information will originate from; online sources, AHK Group, GTK (e.g. possibly automated mineralogy, optical mineralogy, SEM, LA-ICP-MS), Microtrace Solutions USA and Tailorlux Germany. It is possible that other data will be used as the research project advances, but these sources and types and not currently foreseeable. Some of the data sources are subject to approval and copyright permissions being granted.
Funditec	Not applicable.
ULEI	The data to be used to generate new information to be included in the deliverables will have either originated in other organizations (both public and private, that will be identified) or within the Leiden University CML institute.
MO Group	MO Group will receive mineral samples and generate measurement data (e.g. spectra containing characteristic peaks for various elements). The exact origin of the samples will be determined during the project once the suitable candidates (operational mine sites) have been identified.
Spherity	The data used will be shared by the consortium partners.

Table 5 Origin of data generated or re-used by partners

For more information and guidelines see Annex I, 8.4 Data sharing.

2.6 Data utility outside the project

Possible stakeholders are listed in the table below:

Sector	Stakeholders
Projects/Initiatives	EU funded Projects Other grant funded projects Initiatives

Industry	All actors in the raw materials Supply Chain Exploration companies Recyclers Automotive industry Mining, processing and refining companies Relevant manufacturers Producers of recycled intermediates, converters Other industry manufacturers using recycled plastics, metals and minerals Technology and solution providers Online SRM trading platforms Mineral exploration companies Mineral resources companies Mining companies Smelters and refiners Minerals and metals recyclers Mineral commodity traders
Investors	Accelerators Angels Corporates Venture Capital Private Equity Government non-profit Advisor Banks Peer-to-peer investors Mineral exploration investors
EU/MS governments, policy makers and institutions	Ministries Municipalities EU institutions & Regulators Environmental agencies National and regional governments and regulators/authorities International organizations
Research peers: academic sector	Universities Research organizations Standardization and certification bodies Geological institutes
General public	Consumer associations Other
Other: consultancies, NGOs, media	Consultants NGOs Media Law enforcement agencies National or federal police Forensic investigators

Table 6 List of main an sub-groups of stakeholders

For more information and guidelines see Annex I, 8.4 Data sharing.

3 Data Management Policy: FAIR Data
3.1 Making data findable, including provisions for metadata

Digital Object Identifier (DOI)

All open data, publications and open source software produced in MaDiTraCe will be identifiable and findable thanks to a persistent Uniform Resource Locator (URL). Open MaDiTraCe results will be assigned a Digital Object Identifier (DOI) whenever possible to ensure that content is easily and uniquely citable.

Open results that are deposited in the default Open Access repository (Zenodo, see below) will be assigned a DOI automatically and will benefit also from Zenodo's DOI versioning support.

Open results that are deposited in institutional repositories, repositories of scientific publishers or other data and research repositories will be definable at least by a persistent URI. If the institution is a DOI registrant that has an agreement with a DOI registration agency, a DOI will also be assigned.

Whether scientific publications are assigned a unique identifier like DOI, Publisher Item Identifier (PII), International Standard Serial Number (ISSN), etc. depends on the open access strategy (green or gold) chosen by the editors and thus also on the respective scientific publisher and the chosen research repository.

Naming conventions

Files and folders stored in the data repositories will be versioned and structured by using the following name convention:

FileType_FileVersion/Number_MaDiTraCe_PartnerNumber-
PartnerName_YYMMDD.FileExtension (ex. D6.5_V.01_MaDiTraCe_P1-BRGM_210603.doc)

FileType can be:
- D for Deliverable
- DS for DataSet
- F for File (generic, ex. images, table, document).

Zenodo

MaDiTraCe open data will be collected in Zenodo, an open online research data repository. Its repository structure, facilities and management comply with FAIR data principles. Zenodo is an OpenAIRE on which researchers can deposit both publications and data; it provides tools to link them through persistent identifiers and data citations. Zenodo is set up to facilitate the finding, access to, re-use and interoperation of data sets, which are the basic principles that ORD projects must comply with. Zenodo repository is provided by OpenAIRE and hosted by CERN. This catch-all repository enables researchers, scientists, EU projects and institutions to:
- Share research results in a wide variety of formats including text, spreadsheets, audio, video, and images across all fields of science;
- Display their research results and get credited by making the research results citable and integrating them into existing reporting lines to funding agencies like the European Commission;
- Easily access and reuse shared research results;
- Assign a DOI automatically to each result deposited in the repository;
- Integrate their research outputs with the OpenAIRE portal.

A Zenodo community for MaDiTraCe project is available here: https://zenodo.org/communities/maditrace/records?q=&l=list&p=1&s=10&sort=newest

Search keywords

Zenodo allows to perform simple and advanced search queries using keywords. Zenodo also provides a user guide with easy to understand examples. The Data Controllers at each pilot site will be responsible for uploading public datasets that they have generated and to assign specific keywords relevant to these datasets. Dataset specific keywords must describe the content of the dataset. In addition, MaDiTraCe has defined a set of general keywords that should apply to all public datasets, scientific publications and public deliverables, which are the following:

- Critical Raw Materials
- Digital Product Passport
- Traceability
- Chain of custody
- Verifiable credentials
- Blockchain.

Version numbers

Individual file names and datasets will contain version numbers that will be incremented at each revision (V*xyz*). For publications, versioning is in general not necessary.

Zenodo provides DOI versioning of all datasets uploaded to their communities, which allows us to edit and update the uploaded datasets after they have been published. This also allows us to cite specific versions of an upload and cite all versions of an upload.

Metadata

Metadata corresponds to data on the research data themselves. It enables other researchers to find data in an online repository and is therefore essential for the reusability of the dataset. By adding rich and detailed metadata, other researchers can better determine whether the dataset is relevant and useful for their own research. Metadata will be uploaded in a standardised form and be kept separate from the original raw research data.

All metadata will be made available and licensed under CC0, as soon as the different topic specific metadata schemas are established within the project.

The common repository checklist that project partners should use before including one as part of the list of trusted repositories includes the following criteria:

- All datasets that are shared openly on trusted repositories must be available for at least ten years.

- Metadata must remain openly available, even after the datasets have been discarded once the data retention period has expired. This also depends on the arrangement of the repository to warrant functionality over a given period of time.

The following deposition metadata fields are mandatory for MaDiTraCe:
- the terms "European Union (EU)" and "Horizon Europe";
- the name of the action, acronym and grant number;
- the title and description of the deposition metadata,
- the upload type (publication, dataset, software, …),
- the publication date (ISO8601 format, YYYY-MM-DD), and length of embargo period if applicable;
- the creators/authors of the deposition,
- the persistent identifier (DOI),
- the used keywords.

This minimal metadata schema can be extended by arbitrary subjects upon request of the consortium.

For more information and guidelines see Annex I, 8.4 Data Sharing and 8.5 Metadata & documentation.

3.2 Making data openly accessible
3.2.1 Data made openly available as the default

In order to maximise the impact of MaDiTraCe research data, the results are shared within and beyond the consortium. Selected data and results will be shared with the scientific community and other stakeholders through publications in scientific journals and presentations at conferences, as well as through open access data repositories.

The MaDiTraCe project datasets are first stored and organised in a database by the data owners (personal computer, or on the institutional secure server) and on the project database (project website). All data are made available for verification and re-use, unless the task leader can justify why data cannot be made openly accessible. To protect the copyright of the project knowledge, Creative Commons license will be used in some cases.

The MaDiTraCe dataset deliverables are both public (data access policy unrestricted) and they will be accessible by:
- MaDiTraCe project web site
- Partners database
- OpenAIRE
- Zenodo for ORDP data and datasets
- Open access journals.

All data deposited on Zenodo are accessible without restriction for public.

3.2.2 How the data will be made accessible

MaDiTraCe open results will be made accessible according to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon Europe.

Open data: All open results (data, software, scientific publications) of the project will be openly accessible at an appropriate Open Access repository (such as Zenodo) as soon as possible. Specifically, research data needed to validate the results in the scientific publications will be deposited in a data repository at the same time as a publication. Non-public research data will be archived on the repository using a restricted access option.

Scientific publications: Providing open access to peer-reviewed scientific publications can be ensured either by publishing in green or gold open access journals with or without author processing fees. Any scientific publications from MaDiTraCe and the related bibliographic metadata must be made available as open access and announced on the project website (https://www.maditrace.eu) as well as on the OpenAIRE portal (https://www.OpenAIRE.eu/) and on the R&I Participant Portal (https://ec.europa.eu/research/participants). To automate the process of reporting scientific publications and related research data in OpenAIRE, the publication should be deposited in an OpenAIRE-compliant repository, either by the authors of the publication (green open access) or by a scientific publisher (gold open access). This electronic copy of the publication needs to be deposited in an appropriate open access repository in the first place. Additional forms of disseminating open access papers, including academic social

network sites such as ResearchGate (https://www.researchgate.net/) are also possible.. According to the European Research Council's Guidelines on Open Access, "Venues such as Research Gate or Academia.edu that require users to register in order to access content do not count as repositories. The posting of publications on a personal, institutional or project specific webpage or the deposit in a publicly accessible Dropbox account is not sufficient to satisfy the requirements either."

If the chosen repository is not fully OpenAIRE complaint, the publications or data must be linked at https://www.openaire.eu/participate/claim with the respective funding agency (European Commission). Green open access journals or gold open access journals without author processing fees should be preferred to disseminate scientific publications of the MaDiTraCe project. Nevertheless, visibility and prestige of the journal(translated in the Impact Factor), together with the speed of publication, should be considered when choosing a journal for publication of a manuscript. According to the EC recommendation, authors of the publication are encouraged to retain their copyright and grant adequate licences to publishers.

– Green open access (self-archiving)
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, in most cases after its publication in the journal. The journal must grant the researcher the permission to self-archive the final peer-reviewed article, at the latest, 12 months after publication. To find suitable green open access publishers, researchers are encouraged to consult RoMEO (http://sherpa.ac.uk/romeo), a searchable database of publishers' policies regarding the self-archiving of journal articles on the web and in Open Access repositories.

– Gold open access (open access publishing)
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. Author-publishing fees for gold open access journals can be reimbursed within the project period and budget. Some publishers allow the researcher to deposit a copy of the article in a repository, sometimes with an embargo period. To find suitable gold open access publishers, researchers are encouraged to consult the Directory of Open Access Journals (https://doaj.org/), a service that indexes high quality, peer-reviewed open access academic journals that use an appropriate quality control system.

3.2.3 Methods and/or software needed to access the data

Regarding the mere access to open data deposited as data files in a data repository, there are no special methods or software tools needed. The files can be downloaded from the data repository using a standard web browser. The offline viewing, interpreting, processing and editing of data files downloaded from the data repository heavily depends on the type and format of the data.

3.2.4 Data sharing exceptions

1. Copyright and permissions for re-using third-party data sets
Processing and combining input data from many different sources may lead to unclear IPR situations regarding the generated output data, therefore such repurposed data (e.g. model output data) can only be made open if any of the underlying data (e.g. model input data) is open as well.

In that case, potential users must contact the IPR (Intellectual Property Rights) team or the data owner in order to gain access. If necessary, an appropriate IPR procedure (such as non-disclosure agreement - NDA) will be used.

2. Personal data treatment and confidentiality issues

Datasets referring to the quality and quantity of certain elements at risk, such as people and critical infrastructures, are not open by default as their publication may pose privacy, ethical or security risks.

3. Data-driven business model

Data that will be exploited commercially will not be made open.

3.2.5 Specific partner provisions

- **DMT**: Non-sensitive data produced within WP4 will be made available to the public during and after the project. Sensitive data will be made available for partners upon request during the project, due to the commercial interest of DMT as described in the Consortium Agreement.
- **MU Leoben:** Non-sensitive data produced within WP2 will be made available publicly during and after the project. Sensitive data will be made available for partners upon request or, if necessary, only shared encrypted.
- **AHK:** Non-sensitive data produced during the project will be made available to the public during and after the project. Sensitive or commercial data may be made available for partners upon request during the project, this may require encryption or a NDA due to the commercial interest of AHK as described in the Consortium Agreement.

3.2.6 Sensitive data restrictions

Each partner, who requires a restricted access to their datasets, will set up a procedure that will identify persons who require access to these datasets. This identification and access procedure will be described by the respective partner in detail in the next version of the DMP. The RDM of the project will review the procedure and can make recommendations. If applicable, a data access committee, which involves at least one legal specialist in personal and sensitive data, which grants access to the data based on the previously defined procedure.

DMT: Please refer to chapter 3.2.5. Sensitive data will not be made available to the public, due to the commercial interest of DMT. Access by partners will be decided case by case. The project management of DMT (Data Access Committee) will decide who gets access to sensitive data of DMT.

EIT RM: Taking care of stakeholders database. Stakeholders are invited to join the database by link. When collecting information, stakeholders are informed for which purpose the information is used and what kind of information is gathered and they are accepting this by filling the sheet. The information is collected by EIT RM personnel and stored in excel sheet in a SharePoint platform hosted by EIT RawMaterials. The access for excel can be asked from EIT RM for the purpose listed in the information letter behind the link.

ISMC: the management of the newsletter subscription is led by ISMC. By subscribing, members agree to receive emails, with the option to easily unsubscribe at any time. The

collected information is stored in Mailjet, and is accessible only to ISMC for the purpose of sending project updates. Being based in the EU, Mailjet is GDPR compliant.

3.2.7 How access will be provided if there are restrictions on use to data

Where a restriction on open access to research data is necessary, attempts will be made to make data available under controlled conditions to other individual researchers. In the case where restricted or embargoed data is stored in the Zenodo repository, information about the restricted data will be published in the repository, and details of when the data will become available will be included in the metadata. According to the Q&A session "Open Research Data in HE and Zenodo repository", Metadata for open, closed, embargoed and restricted records are always publicly available in Zenodo. Data files and data sets for restricted access records are only visible to their owners and to those the owner grants access. Restricted access allows a researcher to upload a dataset and provide the conditions under which he/she grants access to the data. Researchers wishing to request access must provide a justification for how they fulfil these conditions. The owner of the dataset gets notified for each new request and can decide to either accept or reject the request. If the request is accepted, the requester receives a secret link which usually expires within 1 to 12 months.

For more information and guidelines see Annex I, 8.4 Data sharing.

3.3 Making data interoperable

General data and metadata vocabularies, standards, formats and methodologies will be followed. Nevertheless, there are no generally accepted and standardized field specific data and metadata vocabularies, standards, formats and methodologies. Partners will try to follow community-endorsed interoperability best practice. The majority of the project partners are technically capable of implementing these best practices

Partners will take into consideration OpenAIRE guidelines for online data interoperability: https://guidelines.openaire.eu/en/latest/ and metadata standards according to the Metadata Standards Directory (Engineering, Physical Sciences & Mathematics and General Research Data) guidelines.

LCA data generated in the project will follow main standards in the field of LCA and adapted to main LCA software.

Material fingerprint is a field where there is not, for instance, a data standard.

DPP tasks and developments in WP3 are based on DPP discussions taking place at EU/international level. Interoperability of different DPP systems is one of the main topics to be addressed.

The project is in relation withy UN initiative for interoperability in material traceability.

MaDiTraCe has submitted a number of deliverables that support the harmonization of vocabularies, formats and methodologies since the start of the project. Those deliverables do not represent a real international interoperability document but can be considered as supporting documents:

1. Report on existing tracing and 3D printing technologies (D2.1)
2. Report on identification of related projects and activities (D5.4)
3. Interim State of the Play Report (D1.2)
4. State of the Play Report (D1.3)
5. Analysis and roadmap for implementation of common strategy for the development of material traceability policies within specialisation strategies (D5.5).

3.4 Increase data re-use

The research data management team will advise partners in the project concerning research outputs other than data. The team will develop a process about how to handle the research output generated within the programme. This process will be described in more detail in the next version of the DMP.

For more information and guidelines see Annex I, 8.6 Storage and security

3.5 DMP quality and review process & data inventory

Internal process of quality evaluation and reporting is activated throughout the entire project duration to assess both project data /products and project process (See the D6.4 - Project Quality Plan). An internal peer review is performed for the main project deliverables to guarantee the deliverable is developed with a high level of quality. Each WPL has to submit all the produced documents to another partner assigned as internal reviewer who checks the quality of the documents prepared.

Results data will also be analysed and collected throughout the entire project duration. To this purpose, the Dissemination and Communication Plans (D5.1) will also be filled in by each partner about every six months: it includes the description of articles, papers and scientific publications too. Thus, all research data generated and publications will be analysed and described by using the Data Inventory Table (Annex II), WPLs and partners authors of publications are required to fill in periodically.

Further updating of the DMP will include the updating of online research data repository where data are collected and shared and the data the description of dataset and research data gradually generated and collected.

3.6 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.).

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.

4 Data security
4.1 Data access

The following guidelines will be followed in order to ensure the security of the data:
- Store data in at least two separate locations to avoid loss of data;
- Encrypt data if it is deemed necessary by the participating researchers;
- Limit the use of USB flash drives;
- Label files in a systematically structured way in order to ensure the coherence of the final dataset.

All deliverables and data of MaDiTraCe will be stored and shared in the dedicated Teams repository restricted to the project consortium. Initially, only the Consortium Partners will have access to the repository where dataset and metadata are filed. The protection of data will be ensured through procedures and appropriate technologies (e.g., HTTPS protocol for the encryption of all internet transactions and appropriate European and Internet security standards from ISO, ITU, W3C, IETF and ETSI). If data is kept in a certified repository, then the security standards of that repository will apply.

In the next steps, scientific publications and articles, the dataset deliverables and the final demonstrator research results will be shared through Zenodo and other database to promote the data making FAIR.

4.2 Data preservation

Data preservation refers to data upkeep and maintenance to ensure that the integrity of the data is upheld in the future. This includes properly maintaining the data repository and data backups to ensure the long-term value of the data.

Data backups are expected to occur once a week, though this is subject to change based on the amount of data that will ultimately need to be backed up. Estimating the size of the data is difficult to accomplish at this time since the data that will be initially included in the project has not been fully identified yet. To date, the following datasets have been identified:

- DMT: Standard and audit catalogue documents (small data size, refer to chapter 2.4); Data size per audit < 50 MB per audit (refer to Annex 9); Literature review > 5 GB; all data is stored on TÜV NORD servers to to ensure that the integrity of the data is upheld in the future.
- BRGM: T2.2 and T2.1 related data about lithium in the range of a few 10s of GB. These include stable isotope, Raman-spectroscopy, LIBS and LA-ICP-MS data, as well as XRD and TEM data. Data includes raw data and processed data, related manuscripts and presentations and images from nano to macro scale. All the data is stored in BRGM data storage solution.
- GTK:
  - T2.2: Trace element composition and S isotopes in Ni-Co ore deposits. Excel. Size: < 10 GB

- o T2.3: XCT data - subject: 3D data of mineral concentrate particles, format: Pergeos software .am 3d files, size - approx 4GB per dataset
  - o T2.3: 3D printing data - subject: 3D surface models of tracking particles and 3D printer 'sliced' files for instructing the printer what to print, format: a) .stl 3D surface file, b) Ultimaker.ufp and Formlabs.form files, sizes: 1kb to 1GB, dependent on how many individual particles per print run.
- MUL:
  - o T2.2 related data about graphite in the range of a few 10s of GB. These include stable isotope, Raman-spectroscopy, LIBS and LA-ICP-MS data, as well as XRD and TEM data. Data includes raw data and processed data, related manuscripts and presentations and images from nano to macro scale.

However, MaDiTraCe is designed and intended to have useful data added on a consistent basis, so the size will likely change over time, as well.

The Teams repository hosts part of the global project data and has no limit as to the size of data that can be stored.

The MaDiTraCe data can be downloaded and copied to personal computers without monitoring by each partner having access to the project repository.

5 Ethical aspects
5.1 General Data Protection Regulation (GDPR)

The MaDiTraCe project fully complies with the General Data Protection Regulation (GDPR) laid out in Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC and respects regulations on intellectual property rights (IPR).

Repository and database access will be fully anonymous. If a user wishes to use the Teams of the MaDiTraCe project, they will be required to create an account. The user must provide their email address and grant their consent that their email address be processed and used for account-related communications and management. The user is free to revoke their consent at any time, which would lead to the deletion of their access to the Teams platform.

Between partners having in charge stakeholders' relation and personal data storage (BRGM, MUL, ISMC and LGI), a Joint Data Control Agreement was signed. This document establishes the rules during the project to store and share personal data.

5.2 Sensitive data

Sensitive data is data that is either private or confidential and includes personal user data. The proper management of sensitive data is imperative to maintain the individual privacy and remain in compliance with both EU and international regulations.

In order to ensure proper management of sensitive data, data that is considered sensitive should first be identified. Thus, the main ethical and privacy issues with sensitive data arise from ensuring the data remains private and that proper consent is obtained before the data is shared or published in any way. Sensitive data will be stored privately by partners

handling the data. Measures to protect the privacy of individuals providing sensitive data will be taken in any instance where sensitive data will be collected and published. Furthermore, when partners are meant to deal with sensitive data, they must notify the Consortium and comply with the Joint data controller guidelines (see annex III) and complete the table under annex IV.

6 Allocation of resources

Costs

Costs related to open-access to research data in HE are eligible for reimbursement under the conditions defined in the GA, but also other articles relevant for the cost category chosen. Project beneficiaries will be responsible for applying for reimbursement for costs related to making data accessible to others beyond the consortium.

The costs for making data FAIR includes:
- Fees associated with the publication of scientific articles containing project research data in "Gold" Open access journals. The cost sharing, in case of multiple authors shall be decided among the authors on a case-by-case basis;
- Project Website operation: no cost in addition to standard website maintenance at this stage;
- Data archiving at Zenodo and on other on line data base: free of charge;
- Copyright licensing with Creative Commons: free of charge.

Each partner is responsible for the data they produce. Any fee incurred for Open Access through scientific publication of the data will be the responsibility of the data owner (authors) partner(s).

No immediate costs are anticipated for open data that is stored for long-term preservation in the Zenodo repository. Additional details will be reported, as needed, in future versions of the DMP.

Responsibility

Data management activities concern the whole project and need to be coordinated and monitored both at project and at work package level. Data management is also linked to publication of project results and thus dissemination activities.

The overall responsibility for data management lies with the project coordinator, Daniel Monfort from BRGM.

The members of the Executive Committee (i.e. WPLs) are also responsible of the Data Management of MaDiTraCe dataset and research data.

The Coordinator (project data manager) and the WPLs (WP data managers) are responsible for:
- Developing and implement the data management plan and policy in cooperation with the project management,
- Monitoring data management activities (both collection and publication) and deadlines,
- Monitoring that open results (data and software) are deposited in the default repository or a complementary OpenAIRE-compliant repository and that are linked with MaDiTraCe,
- Providing support and sending reminders to all partners,

- Writing and delivering the data management plan,
- Offering assistance in choosing the right publication path (green or gold open access),
- Offering customized help and further guidance for publishing scientific publications,
- Ensuring that the open access policy of the journal complies with the HE open data requirements before the researcher submits a manuscript,
- Monitoring that green access (self-archiving) publications are deposited in repositories and sending reminders to partners,
- Ensuring that metadata about publications is made available in the R&I Participant Portal (preferably automatically through OpenAIRE) and on the MaDiTraCe website,
- Monitoring that research data related to a publication is made available in repositories and linked to respective publication,
- Checking possible embargo periods and sending reminders to partners,
- Monitoring that publications available in OpenAIRE are properly linked with MaDiTraCe,
- Describing the data (by means of appropriate metadata) or scientific publication in accordance to the MaDiTraCe data management policy (e.g. according to the chosen metadata standard) and with help of the tools (e.g. template, web form, ...) provided by the project.

For more information and guidelines see Annex I, 8.6 Storage and security, 8.8 Roles and Responsibilities, 8.9 Preservation

7 Conclusion

This document describes the main principles and guidelines for Data Management for the MaDiTraCe project. As it is a living document, it will be updated throughout the project lifetime. Further updating of the DMP will include the possible update of online research data repository where data are collected and shared, the standard of data structuration, the description of datasets and research data gradually generated and collected.

Bibliography

Huberman, N., & Pearlmutter, D. (2008). A life-cycle energy analysis of building materials in the Negev desert. Energy and Buildings, 40(5), 837–848. doi:10.1016/j.enbuild.2007.06.002

Bending, R., & Eden, R. J. (1984). UK Energy: Structure, Prospects, and Policies. Cambridge, UK

European Commission. *Horizon Europe Programme. Guidance for the classification of information in research projects.*

European Commission. *Horizon Europe Programme. Guidelines on FAIR Data Management in Horizon Europe.*

European Commission. *Horizon Europe Programme. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon Europe.*

European Union (2020). Intellectual property rights. *https://europa.eu/youreurope/business/running-business/intellectual-property/rights/index_en.htm*

European Parliament (2021). *Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32021R0695*

European Parliament (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:32016R0679*

8 Annex I - Digital Curation Centre (DCC) – Information

This chapter summarizes information provided by the Digital Curation Centre (DCC). The DCC was founded in 2004 to enable research data management the use of data that are findable, accessible, interoperable and reusable (FAIR). The information shared in this chapter can be seen as a guideline for the implementation of a successful Data Management Plan.

8.1 Data Collection

- Outline how the data will be collected and processed. This should cover relevant standards or methods, quality assurance and data organisation.
- Indicate how the data will be organised during the project, mentioning, e.g., naming conventions, version control and folder structures. Consistent, well-ordered research data will be easier to find, understand and reuse.
- Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture, data entry validation, peer review of data or representation with controlled vocabularies.
- See the DataOne Best Practices for data quality.

8.2 Data Description

- Give a summary of the data you will collect or create, noting the content, coverage and data type, e.g., tabular data, survey data, experimental measurements, models, software, audio-visual data, physical samples, etc.
- Consider how your data could complement and integrate with existing data, or whether there are any existing data or methods that you could reuse.
- Indicate which data are of long-term value and should be shared and/or preserved.
- If purchasing or reusing existing data, explain how issues such as copyright and IPR have been addressed. You should aim to minimise any restrictions on the reuse (and subsequent sharing) of third-party data.

8.3 Data volume

- Note what volume of data you will create in MB/GB/TB. Indicate the proportions of raw data, processed data, and other secondary outputs (e.g., reports).
- Consider the implications of data volumes in terms of storage, access and preservation. Do you need to include additional costs?
- Consider whether the scale of the data will pose challenges when sharing or transferring data between sites; if so, how will you address these challenges?

8.4 Data sharing

- How will you share the data e.g. deposit in a data repository, use a secure data service, handle data requests directly or use another mechanism? The methods used

will depend on a number of factors such as the type, size, complexity and sensitivity of the data.

- When will you make the data available? Research funders expect timely release. They typically allow embargoes but not prolonged exclusive use.
- Who will be able to use your data? If you need to restrict access to certain communities or apply data sharing agreements, explain why.
- Consider strategies to minimise restrictions on sharing. These may include anonymising or aggregating data, gaining participant consent for data sharing, gaining copyright permissions, and agreeing a limited embargo period.
- How might your data be reused in other contexts? Where there is potential for reuse, you should use standards and formats that facilitate this, and ensure that appropriate metadata is available online so your data can be discovered. Persistent identifiers should be applied so people can reliably and efficiently find your data. They also help you to track citations and reuse.

8.5 Metadata & documentation

What metadata will be provided to help others identify and discover the data?

- Researchers are strongly encouraged to use community metadata standards where these are in place. The Research Data Alliance offers a <u>Directory of Metadata Standards</u>. Data repositories may also provide guidance about appropriate metadata standards.

- Consider what other documentation is needed to enable reuse. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, any assumptions made, the format and file type of the data and software used to collect and/or process the data.
- Consider how you will capture this information and where it will be recorded, e.g., in a database with links to each item, in a 'readme' text file, in file headers, etc.

8.6 Storage and security

- Describe where the data will be stored and backed up during the course of research activities. This may vary if you are doing fieldwork or working across multiple sites so explain each procedure.
- Identify who will be responsible for backup and how often this will be performed. The use of robust, managed storage with automatic backup, for example, that provided by university IT teams, is preferable. Storing data on laptops, computer hard drives or external storage devices alone is very risky.
- See UK Data Service Guidance on <u>data storage</u> or DataONE Best Practices for <u>storage</u>.
- Also consider data security, particularly if your data is sensitive e.g., detailed personal data, politically sensitive information or trade secrets. Note the main risks and how these will be managed. Also note whether any institutional data security policies are in place.
- Identify any formal standards that you will comply with, e.g., ISO 27001. See the DCC Briefing Paper on Information Security Management - <u>ISO 27000</u> and UK Data Service guidance on <u>data security</u>.

8.7 Ethics and privacy

- Investigators carrying out research involving human participants should request consent to preserve and share the data. Do not just ask for permission to use the data in your study or make unnecessary promises to delete it at the end.
- Consider how you will protect the identity of participants, e.g., via anonymization or using managed access procedures.
- Ethical issues may affect how you store and transfer data, who can see/use it and how long it is kept. You should demonstrate that you are aware of this and have planned accordingly.
- See UK Data Service guidance on consent for data sharing.
- See ICPSR approach to confidentiality and Health Insurance Portability and Accountability Act (HIPAA) regulations for health research.

8.8 Roles and Responsibilities

- Outline the roles and responsibilities for all activities, e.g., data capture, metadata production, data quality, storage and backup, data archiving & data sharing. Individuals should be named where possible.
- For collaborative projects you should explain the coordination of data management responsibilities across partners.
- See UK Data Service guidance on data management roles and responsibilities or DataONE Best Practices: Define roles and assign responsibilities for data management.

8.9 Preservation

- Outline the plans for data sharing and preservation - how long will the data be retained and where will it be archived? Will additional resources be needed to prepare data for deposit or meet any charges from data repositories?
- See the DCC guide: How to appraise and select research data for curation or DataONE Best Practices: Identifying data with long-term value.

9 Annex II – Data sets collected in the project

Data set reference and name: MaDiTraCe – WP1 Organisation in charge: MUL		
Data Type	**Data Standards - Formats**	**Data Generation Software**
Data outputs such as spreadsheet and reports	.docx, .xlsx, .pdf	MS Office
Estimated Data Size	**Data Sharing**	**Storage and Preservation**
1 GB	Non-sensitive data will be accessible to the public during and after the project. However, sensitive data will be made available upon	WP1 folder Teams

Funded by
the European Union

	request as described in the CA.

Name of the data set	Need and gaps in due diligence
Description	• Contains the name and contacts of stakeholders and will be used to establish a continuous process of feedback and acceptance of the MaDiTraCe certification system (This is protected and confidential information). The other part of the information is an inventory of the current state of due diligence initiatives, regulations and frameworks to identify gaps and needs in the manufacturing industry. • List: Policy makers, Raw materials companies, Initiatives and organizations for responsible mining. E.g.: OCDE; IRMA; ICMM; CCCMC; ETI; EITI; GSMA; RMI; EU.
Media Type	Text
Language(s)	English
Use & re-use	Assessment of needs and gaps in due diligence
Size	1 GB
Format/license	.docx, .xlsx, .pdf
Version number	Not relevant

Data set reference and name: MaDiTraCe – WP2 Organisation in charge: BRGM		
Data Type	**Data Standards - Formats**	**Data Generation Software**
Analytical data (trace element composition, isotopic compositions, etc.)	.xls, .csv, .txt	MS Office
Estimated Data Size	**Data Sharing**	**Storage and Preservation**
< 50 MB per folder	Zenodo	Zenodo and internal BRGM database

Name of the data set	Analytical results on the different project samples
Description	• These are analytical results on different samples for the different commodities. These results will be used to develop the geochemical traceability of supply chains. • The results may be published in a scientific journal.
Media Type	Numerical data
Language(s)	English
Use & re-use	Geochemical traceability development / no re-use planned
Size	Data sizes will be small mainly xls folders based < 50 MB
Format/license	.xls, .csv, .txt
Version number	1.0, regularly updated

Data set reference and name: MaDiTraCe – WP3 Organisation in charge: CEA		
Data Type	**Data Standards - Formats**	**Data Generation Software**
• Questionnaires for State of play analysis. No	.xls, .xml, .docx, .csv	• Not applicable. Various sources are used (please refer to the ULEI

sensitive data. Examples of Flow Models. • Supply chain quantification for the supply chain mapping (T3.1) • Life cycle analysis of selected materials/ products (T3.2).		contributions on the data management plan, Deliverable 6.5, for more information). • Excel • Brightway
Estimated Data Size	**Data Sharing**	**Storage and Preservation**
<1G	• Non-sensitive data used for demonstration purpose will be accessible to the public during and after the project in concertation with use cases providers in other work packages. Use Cases Examples will be used for potential publication. • No real data set will be produced in this WP. • Relevant data will be shared in the reports for T3.1 and T3.2, scientific publications.	• WP3 folder Teams • Journals

Name of the data set	• **Model flows and related examples of decentralized services** • **Supply chain flows for selected materials** • **LCA results for selected materials**
Description	• A brief, easy to understand description of what the dataset contains and what it will be used for in the project. • A list of institutions to whom the data set could be useful outside the project. • Whether the dataset has been/will be used for a scientific publication (if yes, brief details about the content and journal). • If the dataset is collected, a brief description of its origin and how it was collected will be provided. • Material flows of the supply chains of the selected materials, including origin and destination of each flow, for primary and secondary value chains. • LCA results for the selected material.
Media Type	Numerical data, .xls
Language(s)	English
Use & re-use	• Foreseen use of the resource for which it has been produced. • Usage for demonstration and communication purpose. These results will be used to identify leverage points for

	traceability technologies, and to perform a state of practices analysis regarding control methods and tracing solutions used at different check points.
	• Usage for demonstration and communication purpose. The LCA results will be used to quantitatively describe the production processes of the selected materials from the mining to the materials to the semi-products/ materials that are used in batteries (Li, Co, graphite) and electromotors (REEs/magnets). This description can then be used to better understand how the physical/chemical involved will influence the traceability of the materials, as stated in the project proposal.
Size	Size of the resource with regard to a specific size unit measurement in the form of a number.
Format/license	The format in which the data will be available (e.g. .xls, .csv, .txt) will be provided. Mostly questionnaires and non-sensitive data for demonstration purpose.
Version number	1.0, regularly updated

Data set reference and name: MaDiTraCe – WP4 Organisation in charge: DMT		
Data Type	**Data Standards - Formats**	**Data Generation Software**
Standard documents and audit catalogues / presentations	.xlsx, .docx, .pptx	MS Office
Estimated Data Size	**Data Sharing**	**Storage and Preservation**
< 50 MB per Audit	Non-sensitive data produced within WP 4 will be made available for public during and after the project. Sensitive data will be made available for partners upon request during the project, due to the commercial interest of DMT as described in the CA.	At TUEV NORD server

Name of the data set	Audit results
Description	Audit results out of the audit piloting of CERA 4in1 standards to validate the functionality and feasibility of the CERA 4in1 scheme. The outcome of every audit typically consists of the report and the filled in audit check list including evidences (e.g. documents, photographs). This information is strictly confidential and cannot be used outside the project or for scientific publication.
Media Type	Data will be stored on the TUEV NORD server
Language(s)	Preferably English, but depends on the company to be audited
Use & re-use	Standard validation / no re-use planned

Size	Data sizes will be small mainly text based (standard documents and audit catalogues) as well as evidences < 50 MB per Audit
Format/license	Various MS Office / image formats
Version number	1.0, regularly updated

Data set reference and name: MaDiTraCe – WP5 Organisation in charge: ISMC		
Data Type	**Data Standards - Formats**	**Data Generation Software**
Analytical data, reports, documents, presentations	.docx, .xlsx, .pdf, .ppt, .png	MS Office
Estimated Data Size	**Data Sharing**	**Storage and Preservation**
1 GB max	Non-sensitive data will be accessible to the public during and after the project. However, sensitive data will be made available upon request as described in the CA.	WP5 folder Teams

Name of the data set	Communication, dissemination, exploitation & business case dataset
Description	• The MaDiTraCe C&D dataset will contain the data about C&D actions, supply chain traceability and scientific results compiled throughout the project. These data are expected to be useful to partners and WP leaders for the assesment of WP5 progress. • Multiple audiences beyond the project's own community, incl. media and the broad public, also those that may take an interest in the (potential) USE of the results (e.g. the scientific community, industrial partner, policymakers) or interest in different material origins. • MaDiTraCe website for public reports and OA online repositories such as Zenodo to make scientific publications accessible.
Media Type	Video, image, text, numerical data, n-grams, infographics
Language(s)	English
Use & re-use	Communication and Dissemination uses
Size	1 GB max
Format/license	.docx, .xlsx, .pdf, .ppt
Version number	Not relevant

Data set reference and name: MaDiTraCe – WP6 Organisation in charge: BRGM		
Data Type	**Data Standards - Formats**	**Data Generation Software**
Reports, spreadsheets	docx, xlsx, pdf	MS Office
Estimated Data Size	**Data Sharing**	**Storage and Preservation**
>1GB	Data will only be shared with Consortium	Teams repository of the project

Name of the data set	Consortium members contacts, projects management tool
Description	Spreadsheets that will help track contacts withing the consortium and monitor the status of deliverables and milestones within the project.
Media Type	Spreadsheets
Language(s)	English
Use & re-use	Facilitate the communication between members and monitoring of the project
Size	Less than 1GB
Format/license	.xls, .txt
Version number	Not relevant

10 Annex III – Data sets could be made openly accessible

Data producer	Brief description of dataset	Foreseen use & re-use	Possibility to share the data beyond the consortium
DMT	Data collected from public sources (e.g. for standard development)	Standard development / no re-use planned	Tbd on a case by case basis

11 Annex IV - Management of sensitive data Guidelines

These guidelines are aligned to the European Union regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95.46.EC (General Data Protection Regulation). Referred to as from now onward as GDPR.

11.1 Definitions

"Personal Data"
Personal Data means any information relating to an identified or identifiable natural person ("**Data Subject**"), an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

"Data types used in the [MaDiTraCe] project" shall be confined to the following main categories of Data Subject:
The names, e-mail and mail addresses, telephone numbers, job titles, job profiles, employer, classification of the knowledge system (type of research institution, type of hybrid

knowledge institution), classification of the economic system (type of industry sector, type of industry umbrella organisation, type of commerce, type of technical sector), classification of the political system (type of executive system, type of legislative system), classification of civil society (type of civil society engagement, type of civil society funding institutions, type of community, individuals), classification of hidden actors (type of informal actors, type of criminal actors), geographical information.

"Controller, Data Controller, Processor, Data Processor, Data Subject" as set out in the relevant Data Protection Laws in force at the time.

"Data Discloser"
Data Discloser means a Party transferring the Personal Data to the Data Receiver. Any of the Parties may be a Data Discloser.

"Data Receiver"
Data Receiver means a Party receiving the Personal Data from the Data Discloser. Any of the Parties may be a Data Receiver.

"Permitted Recipients"
Permitted Recipients means the Parties to this Agreement, the directors and employees of each Party, and any third parties that might be engaged to perform obligations in connection with the Project.

"Process/Processing"
Process/Processing means any operation or set of operations which is performed on Personal Data or on sets of Personal Data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.

"Profiling"*
Profiling means any form of automated processing of Personal Data consisting of the use of Personal Data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour or movements.

11.2 Guidelines

Each partner is responsible for the Data Subjects with whom the individual partner enters into relationship, including the responsibility:

- to inform the „Data Subject" of the Processing of Personal Data and the rights of the Data Subject;
- to respond to Data Subjects when contacted by them in accordance with the GDPR;
- to ensure that the necessary preconditions exist for the Processing of Personal Data, including the obtaining of consent; and
- that data are erased when they are no longer necessary.

The nature of the relationship to the Data Subjects is explained in more detail in Annex 1.

If a Data Subject contacts one of the Partners in relation to MaDiTraCe and GDPR, that partner has to immediately forward the request to the other partners. The partners will assist each other in responding to requests and provide the necessary information on the measures taken in accordance with Articles 15 to 22 GDPR.

11.2.1 Information to be provided according to Articles 13 and 14 GDPR

The Parties will provide the Data Subject with all information required under Articles 13 and 14 GDPR by means of a data protection notice on the MaDiTraCe project website.

The Data Subject will be informed on the subject of the MaDiTraCe project and advised that a participation is voluntary and does not involve any risks, costs or similar unpleasant consequences for them. Participation in the project and associated data collection is based on the consent of the Data Subject.

11.2.2 Obtaining Personal Data from other sources

The partner who obtains specific data from sources other than the Data Subject itself is responsible for informing the Data Subject accordingly and requesting separate permission for data use.

11.2.3 Principles and authority to process data

The partner who obtains specific Personal Data is responsible for ensuring:

a) that there is a valid legal ground for Processing and for documenting this to both supervisory authorities and the Data Subject;
b) that the data are not further processed in a manner that is incompatible with the purposes for which they were originally collected.

All partners shall include the Processing operations in their records in accordance with Article 30 paragraph 1 GDPR (as far as applicable to the Parties).

If a data protection impact assessment pursuant to Article 35 GDPR is required, it will be carried out jointly by the partners.

12 Annex V – Data sets to remain confidential

Data producer	Brief description of dataset and means for processing	Purpose of processing	Format	Lead partner	Origin
DMT	Pilot audit results	Validation of standards	Docx / xlsx	DMT	Audited companies

=== End of the document ===