Multi Miner

www.multiminer.eu

@company/multiminer

@MultiMiner_EU

Grant Agreement No: 101091374

Project Start Date: 1st January 2023

Project Duration: 42 Months

Deliverable No 5.4
Data Management Plan

Grant agreement	**101091374**
Project title	Multi-Source and Multi-Scale Earth Observation and Novel Machine Learning for Mineral Exploration and Mine Site Monitoring
Project acronym	MultiMiner
Project coordinator	Dr. Maarit Middleton, GTK
Project start date	Jan 1, 2023
Project duration	42 months
Related work packages	WP 5 – Communication, dissemination and exploitation
Related task(s)	Task 5.6 Data Management and Intellectual Property
Lead organization	GTK
Contributing partners	All
Submission date	May 31, 2023
File name	D5_4_MultiMiner_Data_Management_Plan_V1_3.docx
Organisation responsible for deliverable	GTK
Author name(s)	Dr. Kati Laakso
Date	May 31, 2023
DMP version	1.3
Status	Complete
Dissemination Level	Public

Revision History

Version	Date	Modified By	Comments
V1	09.02.2023	Kati Laakso	First version
V1.1	05.03.2023	Maarit Middleton	Minor comments
V1.2	13.03.2023	Matthieu Molinier	Minor comments
V1.3	28.05.2023	Kati Laakso	Made minor changes based on comments from the consortium.

Disclaimer

The contents of this publication are the sole responsibility of its author and do not necessarily reflect the opinion of the European Union.

Table of Contents

List of Tables

List of Acronyms

Table 1 A list of acronyms used in this document

Abbreviation	Meaning
CRM	Critical Raw Materials
DMP	Data Management Plan
EC	European Commission
EO	Earth Observation
RDM	Research Data Management

Vocabulary

Table 2 Document vocabulary

Term	Meaning
DCAT	DCAT is an RDF (Resource Description Framework) vocabulary designed to facilitate interoperability between data catalogs published on the Web ([link](#))
Dublin Core	"The Dublin Core", also known as the Dublin Core Metadata Element Set, is a set of fifteen "core" elements (properties) for describing resources ([link](#))
JSON-LD	JSON is a useful data serialization and messaging format ([link](#))
MARCXML	MARCXML is a Document Type Definition (DTD) describing the MARC 21 format in XML. MARCXML is used in many applications at the Library of Congress and in OCLC WorldCat and was designed to assist the evolution of bibliographic formats towards XML, while maintaining compatibility with existing bibliographic data ([link](#))
OAI-PMH	The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a low-barrier mechanism for repository interoperability ([link](#))
Research output	Data, software, algorithms, apps, protocols, models, electronic notebooks and workflows
REST API	A REST API is an application programming interface that stands for Representative State Transfer ([link](#)).

1. Executive Summary

The Multi-source and Multi-scale Earth observation and Novel Machine Learning Methods for Mineral Exploration and Mine Site Monitoring (MultiMiner) project develops novel data processing algorithms for cost-effective utilization of Earth Observation (EO) technologies for mineral exploration and mine site monitoring. The project focuses on new EO-based exploration technologies for critical raw materials (CRM) to increase the probability of finding new sources within EU thereby strengthening the EU autonomy in the area of raw materials.

The importance of good data management practices is emphasized in data-intensive projects such as MultiMiner. During the project, large and heterogenous datasets will be acquired and analyzed. The project comprises data acquired in different scales and by different technologies. The number of the consortium partners (13) highlights the importance of data harmonization, i.e. efforts to collect and distribute data using collectively agreed methods, formats and metadata descriptions.

MultiMiner generates and collects a range of data types, most of which are instrument data. Examples of such data include, but are not limited to, point spectral data, radar satellite data and laboratory-collected hyperspectral image data. These data are stored in internationally accepted file formats for their effective sharing, re-use, and preservation. The total estimated size of the data is in the order of terabytes. MultiMiner is reusing publicly available satellite data (e.g. EnMAP, Sentinel) and existing field data from participating test sites. Should any personal data be collected during the project (e.g., stakeholder surveys), all such data are stored in line with the requirements of the General Data Protection Regulation (GDPR).

By default, all data, software and algorithms needed to reproduce and facilitate the reuse of the results of MultiMiner are made publicly available using Zenodo and GitHub. These are repositories that provide persistent identifiers (e.g., DOIs), generate formal citations and whose metadata and data are registered and searchable by search engines to ensure the findability of the data. The MultiMiner consortium shares any data as soon as possible and usually immediately after the data production and quality controls. Here, we follow the principle "as open as possible, as closed as necessary" and should the immediate release of the data be delayed, or the data restricted e.g., due to IP considerations, it is always ensured that the metadata of the data are immediately and publicly available. Only such trusted data repositories are used to share data which provide harvestable and public machine-accessible interfaces to metadata, data files, dataset terms, licenses and version information, making the data findable and accessible.

When possible, all data are prepared such as to include qualified references to other data, maximizing the interoperability of the data. Data are carefully documented at all project phases (data acquisition, data analysis, data sharing) and metadata are compiled using such broad cross-domain and domain-specific metadata schemas (e.g., DataCite, Dublin Core, ISO 19115) and controlled vocabularies (e.g., ISO 3166) which maximize the reusability and interoperability of the data. The metadata of the data always include a detailed provenance, thus including such items as the names of the authors, in addition to information about how the data was collected and processed. By default, MultiMiner shares all its data using clear and permissive licenses to maximize data reusability. By default, all metadata and data are shared using the Creative Commons Public Domain Dedication CC0 and software, algorithms and apps generated in the project are shared using permissive software licenses such as the MIT license. Access to physical samples is given upon request and at the same time as the associated data, following the same principles of

openness. Only if data sharing is against the beneficiary's legitimate interests, including, but not limited to, commercial exploitation of the research outputs, or if the consortium is reusing data already shared by other organizations (e.g., the satellite data) data is not shared or access to it is restricted. All data is stored in the secure servers of the project partners during the project and a minimum of 10 years after the end of the project. Cloud services that are deemed safe are used to share the data between the said partners.

While the Coordinator takes a leading role in data management, it is the responsibility of the individual partner organizations to ensure that the data they generate in the project is managed in accordance with the DMP and meets appropriate national and international legal and ethical standards. Data management activities are adequately resourced and implemented, and all partners allocate a minimum of 5% of their time for research data management. The DMP is treated as a living document and is updated throughout the project, and we publish its different versions in publishing platforms such as Research Ideas and Outcomes RIO journal. Examples of situations where the DMP would be updated include, but are not limited to: new datasets, changes to data storage or data sharing practices, changes in the level of openness of the data (fully open, partially open or fully closed). A new version of the DMP will be created and shared any time major changes take place.

2. Introduction

2.1 Project summary

The Multi-source and Multi-scale Earth observation and Novel Machine Learning Methods for Mineral Exploration and Mine Site Monitoring (MultiMiner) project develops novel data processing algorithms for cost-effective utilization of Earth Observation (EO) technologies for mineral exploration and mine site monitoring. MultiMiner unlocks the potential of EO data, including Copernicus, commercial satellites, upcoming missions, airborne and low altitude as well as in situ data, to support the entire mining life cycle including mineral exploration, operational, closure and post-closure stages. This is achieved by creating generic but highly innovative machine learning solutions which do not require any or only little ground truth data. The project focuses on new EO based exploration technologies for critical raw materials (CRM) to increase the probability of finding new sources within EU thereby strengthening the EU autonomy in the area of raw materials. MultiMiner EO based exploration solutions have extremely low environmental impact, and are thus socially acceptable, economically efficient and improve safety. The project's solutions for mine site monitoring increase the transparency of mining operations as environmental impacts can be detected as early as possible and digital information of the currently unexploitable raw materials can be stored for future generations. The applicability of the developed algorithms is demonstrated in 4 European test sites. MultiMiner is a pan-European consortium consisting of 12 partners and 1 associated partner from research institutes, academia, consulting businesses and mining industry with interdisciplinary backgrounds in geology, remote sensing and machine learning. The members come from six EU member states which represent mining regions across Europe with diverse geology with evident potential for various types of CRM resources and thousands of operational and closed mines.

The strategic aim of MultiMiner is to provide Europe with novel scalable, robust and integrated mineral exploration and mine site monitoring solutions based primarily on multi-source EO data, facilitating discovery of critical raw materials (CRM) and their safe and environmentally sustainable exploitation in Europe. To enable the transition towards the MultiMiner strategy, the following two strategic objectives

(SO) have been defined to address the expected outcomes of the call HORIZON-CL4-2022-RESILIENCE-01-08:

(SO1) To develop scalable methods combining extensive Earth Observation (EO) data at multiple spectral, spatial and temporal resolutions, and making the most efficient use of scarcely available in situ measurements, while improving the efficiency and timeliness of EO-assisted critical raw material exploration, and monitoring of mine operations and environmental impacts.

(SO2) To demonstrate the added-value of the novel products and services covering the whole mining life cycle across Europe, and providing mining stakeholders and EO value-adding industry with pathways contributing to increasing access to critical raw materials in Europe in a sustainable way.

In order to achieve the strategic objectives (SO1) and (SO2), we have specified four general objectives (GO1) to (GO4), each operationalized by 3 to 4 specific objectives:

(GO1) To develop scalable and automated approaches for mineral exploration based on multi-source EO data and sparse in situ data, focused on mineral deposits hosting CRMs across EU.

(GO2) To leverage novel EO data analysis methods to make the most of scarcely available in situ data for timely mine site monitoring, reducing both disruptions to mining activities and environmental impacts.

(GO3) To demonstrate novel exploration and monitoring methods for the whole mining life cycle in 4 test sites across Europe, emphasizing their potential to increase access to critical raw materials across Europe.

(GO4) To share the innovative MultiMiner methods broadly with European value-adding industry to stimulate further research and developments, and ensure their exploitation by the European mining industry.

The project comprises the Consortium Partners listed in Table 3.

Table 3 The Consortium Partners of the MultiMiner project.

Name	Short name	Country	
GEOLOGIAN TUTKIMUSKESKUS	GTK	Finland	
TEKNOLOGIAN TUTKIMUSKESKUS VTT OY	VTT	Finland	
YARA SUOMI OY	YARA	Finland	
ELLINIKI ARCHI GEOLOGIKON KAI METALLEFTIKON EREVNON	HSGME	Greece	

Name	Short name	Country
FONDATION EUROPEENNE DE LA SCIENCE	ESF	France
CESKA GEOLOGICKA SLUZBA	CGS	Czechia
MONTANUNIVERSITAET LEOBEN	MUL	Austria
BUNDESANSTALT FUER GEOWISSENSCHAFTEN UND ROHSTOFFE	BGR	Germany
GEOSPHERE AUSTRIA	Geosphere Austria	Austria
HELLAS GOLD S.A	HG	Greece
EFTAS FERNERKUNDUNG TECHNOLOGIETRANSFER GMBH	EFTAS	Germany
VEITSCH-RADEX GMBH & CO OG	RHI	Austria
TECHNISCHE UNIVERSITAET MUENCHEN	TUM	Germany

2.2 Purpose of the Data Management Plan

The objective of this Data Management Plan (DMP) is to act as guidance for the Consortium Partners in the application of research data management (RDM) practices. The document covers different aspects of RDM, namely, data types, data sharing, data re-use, allocation of resources and data security, in addition to ethical, legal and societal aspects of the data.

The DMP is a living document that will be updated whenever a need arises. Examples of situations where the DMP would be updated include, but are not limited to: new datasets, changes to data storage or data sharing practices, changes in the level of openness of individual datasets. In the MultiMiner project datasets can be fully open, partially open or fully closed. A new version of the DMP will be released and shared each time **significant changes** take place. The different versions of the DMP will be shared in the

CORDIS system of the European Commission as well as in the Research Ideas and Outcomes (RIO) journal (LINK).

2.3 Relation to other project documents and deliverables

The Consortium Agreement and **Grant Agreement** have elements in them that broadly fall into the domain of a DMP. These elements include: confidentiality, intellectual property rights, patents, dissemination of the results and access rights to the results. In the event of discrepancy between the documents, the DMP is overruled by the Consortium Agreement, including its annexes and possible addendums.

Deliverable D4.1 (Field work protocols for in situ data collection and review of accuracy assessment methods) will describe in detail the metadata schemas, requirements and protocols used to acquire data either in the field or in a laboratory.

Open access publishing and RDM fall into the broader domain of open science. Open Access practices are exclusively discussed in the MultiMiner **Project Handbook**.

Deliverable D1.2 (Ethics Requirements): any questions related to ethics are described in D1.2. In this document, they are discussed only shortly in section 3.7.

2.4 Roles and responsibilites

Following good RDM practices, outlined in this document, is the responsibility of each Consortium Partner. GTK is the Coordinator of the project, but also one of the Consortium Partners, depending on the context.

The first version of the DMP is drafted by the Coordinator in collaboration with the other Consortium Partners (Table 4). After the first version, it is the responsibility of each project partner to notify the Coordinator if updates are needed.

Table 4 The distribution of responsibilities in the MultiMiner project.

Consortium Partner	Coordinator
Ensures that the DMP is in agreement with any relevant institutional policies (e.g. data management handbooks).	**Drafts** the first version of the DMP together with the Consortium Partners.
Informs the Coordinator about any relevant changes to institutional open science and RDM guidelines.	**Updates** the DMP based on feedback from the Consortium Partners. **Shares** the DMP in CORDIS and RIO whenever there are relevant updates.
Informs the Coordinator about any research outputs at least 21 days prior to sharing them.	**Approves** the sharing of datasets and other research outputs together with the other Management Committee members (see the Project Handbook for details). This will be done as soon as possible and no later than within 21 days of the request.
Shares their own research outputs in recommended data repositories.	**Ensures** that the list of datasets and project outputs, listed in sections 4.1 and 4.2, are always up-to-date.

2.5 Licensing

The licensing policy of the MultiMiner project follows the guidance of the EC, as outlined in the Annotated Grant Agreement and Programme Guide documents. As listed in Table 5, the CC0 license will be used for data and metadata, whereas for publications, the CC BY 4.0-license will be used. However, open access publishing is beyond the scope of this DMP document and will not be discussed further. Open Access practices are exclusively discussed in the MultiMiner **Project Handbook**.

As the EC does not impose software licensing obligations, permissive licenses, such as the MIT license, will be used to license software and codes in the MultiMiner project.

Table 5 The licenses used in the MultiMiner project.

License	Explanation	Used for
CC0 1.0 Universal Public Domain Dedication	"No copyright. The person who associated a work with this deed has dedicated the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law. You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission".	Data and metadata
CC BY Attribution 4.0 International	"Free to Share — copy and redistribute the material in any medium or format Adapt — remix, transform, and build upon the material for any purpose, even commercially, under the following terms: Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits". (https://creativecommons.org/licenses/by/4.0/)	Publications (open access publishing will be discussed in more detail in the Project Handbook).
The MIT License	"Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to	Software and Codes

| | permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software". (opensource.org/licenses/MIT)

Software licensing will be considered on a case-by-case basis and any potential patenting requirements will be taken into account. | |

2.6 File and folder naming conventions

The project uses a self-explanatory folder structure to safeguard its data. Each of the four study areas has its own structure. The exact content of the site folders depends on the data that will be acquired in them. The folder structure is aimed for internal sharing (within the consortium) and storing of the data, hence, it is not used to share data with third parties. Initially (the first 7 months of the project) the data are stored in the Nextcloud online collaboration platform. After this initial period, the data will be stored in Azure cloud services for a minimum of 10 years or for as long as the data have scientific value. This switch from one platform to another reflects an IT policy change of the Coordinator.

Within the cloud services, file names follow an agreed consistent format, which comprises the following elements (modified from https://libguides.princeton.edu/c.php?g=102546&p=930626):

- File names should be relatively short but descriptive (<35 characters)
- Special characters and spaces are not allowed.
- Underscores will be used instead of periods or spaces or slashes.
- Use date format ISO 8601: YYYYMMDD.

An example on a filename would be MM_<date>_<site_id><task_id><descriptor>
   where
   MM=MultiMiner (an acronym of the project name)
   site_id=S=Siilinjärvi, H=Hochfilzen, K=Kallyntiri, C=Chalkidiki etc.
   task_id=the identifier of the task as given in the DoA (e.g. T4.1).
   descriptor: freeform description of the data, for instance, PRISMA (the satellite name)

Examples on files names:
1. MM_20230615_S_T43_ASD_Jaakonlampi
2. MM_20240703_H_T32_Sentinel

The above file naming conventions will be used for data generated within the project unless it would require unreasonable efforts, for instance, in the case of large quantities of field data, or with Earth Observation imagery that requires another naming convention for compatibility with data processing libraries.

3 The Data Management Plan

3.1 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 will use re-existing data as detailed in Table 6. Existing satellite data are at the core of the MultiMiner project. These data form the basis for all data analysis, and tool/methodology development conducted in the project. Another re-used set of data that are extremely important to the project are existing data from the mine sites. These data include, but are not limited to, lithology descriptions, ground-based radar data, dust and water quality and vegetation inventory.
What types and formats of data will the project generate or re-use?
Listed in Table 6.
What is the purpose of the data generation or re-use and its relation to the objectives of the project?
Related to the project goals (GO1-GO4), listed in Section 2.1, the MultiMiner project will, on one hand, use existing EO data and on the other hand, acquire drone- and manned aircraft-borne data to develop mineral exploration and mine site monitoring tools (GO1-GO2). In this work, existing data (e.g. lithogeochemistry, lithological information) from the mine sites will be used. The project will also acquire new data in the field and in a laboratory to demonstrate the effectiveness of the thus acquired tools (GO3). The data acquired in the project will be shared with the project stakeholders to maximize the impact of the research outputs of the project within the European mining industry (GO4).
What is the expected size of the data that you intend to generate or re-use?
Listed in Table 6.
What is the origin/provenance of the data, either generated or re-used?
Listed in Table 6.
To whom might your data be useful ('data utility'), outside your project?
The data might be useful to the different stakeholders of the project, namely the mining industry, consulting companies, geological surveys, governmental organizations, non-governmental organizations, EO data providers, EO value-adding industry, the scientific and R&D communities, policy / decision makers and local authorities.

3.2 Fair data
3.2.1 Making data findable, including provisions for metadata
Will data be identified by a persistent identifier?
The recommended data repository of the project is Zenodo (https://zenodo.org) which provides DOIs for every published record. For sharing software and codes, GitHub (https://github.com/) is

3.2 Fair data
the recommended repository. GitHub does not, by default, provide DOIs (or other persistent identifiers) to the data, but one will be obtained for each published code or software through Zenodo as explained in <u>these instructions</u>.
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.
Data acquisition: The project has its own metadata schema that will be used for data acquired in the field or in a laboratory. The schema will be described in full in deliverable D4.1 (Field work protocols for in situ data collection and review of accuracy assessment methods). **Data analysis:** Any data analysis protocols will be described in detail in README documents that will be shared with the data. In addition, any protocols related to data acquisition or data analysis will be shared in the <u>protocols.io</u> service if they are considered useful to the wider research community. **Data sharing:** Zenodo's metadata are compliant with <u>DataCite's Metadata Schema</u>. Furthermore, the data acquisition and data analysis protocols of the project will be shared as README files and protocols.io descriptions as discussed above.
Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?
Yes, suitable search keywords will be provided for each dataset that are shared in the trusted repositories (see above for details) using the dedicated keywords-field of the user interface. The following standard keywords will always be used when sharing data or software/codes: *MultiMiner* and either *mineral prospectivity mapping* or *mine site monitoring*, or both, depending on the content of the research output. The non-standard keywords will be decided based on the content of the research output.
Will metadata be offered in such a way that it can be harvested and indexed?
The metadata of the Zenodo data repository are harvestable using the OAI-PMH protocol. Metadata are also retrievable through a public REST API. Metadata are publicly accessible and licensed under public domain. No authorization is ever necessary to retrieve it.   "Metadata of each record is indexed and searchable directly in Zenodo's search engine immediately after publishing".   "Metadata of each record is sent to DataCite servers during DOI registration and indexed there".   (source: <u>https://about.zenodo.org/principles/</u>) GitHub also has a REST API, in addition to a GraphQL API for retrieving the contents of the repository (<u>https://docs.github.com/en/rest?apiVersion=2022-11-28</u>, <u>https://docs.github.com/en/graphql</u>).
3.2.2 Making data accessible
Will the data be deposited in a trusted repository?
By default, all research outputs generated in the MultiMiner project will be deposited in trusted repositories. More specifically, the data generated in the project will be upload in Zenodo and any

3.2 Fair data
algorithms and software generated in the project will be shared in GitHub, as discussed in section 3.2.1. By default, all data, software and codes will be made available upon publication such that the dataset will be linked to the publication, when possible, and the publication will be linked to the associated datasets. The data, software and codes will be made, by default, available immediately and without embargoes.
Have you explored appropriate arrangements with the identified repository where your data will be deposited?
A community will be created in Zenodo to keep all the datasets of the project together. In GitHub, software and algorithm research outputs will be collected under a project/organization.
Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?
Each dataset in Zenodo is given a DOI (a persistent identifier). These DOIs resolve to a webpage that displays both the metadata and the data, in addition to licensing, community information and administrative information (the number of views etc.).
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.
All data, metadata, software and codes created by the MultiMiner project will be made, by default, publicly available. Confidential data will not be made publicly accessible without permission from the owners of the said data. Even in the case of confidential data, the metadata of the data will be made openly public using the CC0 license. Confidential data are listed in Table 6. If at later stages of the project there should be datasets, codes/software or other research outputs that will be restricted temporarily or permanently, those dataset(s) and the reason for restricting their availability will be given in this document. In case the actual data should be restricted temporarily or permanently, the metadata of the data will nevertheless be published immediately using the CC0 license (no rights reserved).
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.
By default, the MultiMiner project will publish all of its data and other research outputs at the time of publication (at the latest). Prior to publishing any research outputs, the project personnel are encouraged to actively consider the possibility of patenting their innovations. Should it be observed that there is potential to obtain patents to any research outputs, such research outputs will not be openly shared, but instead, advise will be seeked from institutional specialists on patenting. Should it be later decided that there are no avenues for patenting, the research output(s) will be immediately published. In any case, the metadata of the data will be published immediately using the CC0 license (no rights reserved).
Will the data be accessible through a free and standardized access protocol?
Yes, the OAI-PMH and REST protocols of Zenodo and GitHub are open, free and universal protocols for information retrieval on the web.

3.2 Fair data

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

In the case of confidential data that cannot be openly published, the metadata of the data will be immediately published nevertheless. The confidential data will be kept in the secure servers of the Consortium Partners for a minimum of 10 years or as long as the data is considered to have scientific value. A detailed listing of confidential data is given in Table 6.

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

The project data that has not yet been openly shared or which is confidential is only available to the consortium members in safe server spaces. Outsiders cannot access any such data.

Data shared in Zenodo and software/codes shared in GitHub are freely available and the project does not track the identities of the people downloading them.

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

No, there is no need for such a committee, because the MultiMiner project does not collect personal data for any research purposes and thus, will not under any circumstances be sharing personal/sensitive data.

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?

Yes, the metadata of the data will be licensed under the CC0 license. By default, the metadata will accompany the data that it is associated with, and hence it is always clear to the potential users how to access the data (or other research outputs).

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

The data will remain available and findable for as long as it has scientific value. Data, software codes, and their associated metadata will remain available for the lifetime of the data repository. In the case of Zenodo, the lifetime is currently defined as a minimum of 20 years (https://about.zenodo.org/principles/). In the MultiMiner project, metadata will always accompany the data it describes and hence it is not likely that at any point the metadata would remain in trusted repositories without the data.

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)?

If specific software is needed to either access or read the data produced in the MultiMiner project, a reference of the said software will be added to the metadata descriptions of the data. By default, all software and algorithms generated by the project will be openly shared using a permissive license (see section 2.5).

3.2.3 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?

Data are carefully documented at all project phases (data acquisition, data analysis, data sharing) and metadata are compiled using such broad cross-domain and domain-specific metadata schemas

3.2 Fair data

(e.g., DataCite, Dublin Core, ISO 19115) and controlled vocabularies (e.g., ISO 3166) which maximize the reusability and interoperability of the data.

Data sharing: Zenodo's metadata is compliant with DataCite's Metadata Schema minimum and recommended terms (https://about.zenodo.org/principles/). Zenodo's metadata can be exported to Dublin Core and MARCXML formats. The following metadata formats are provided by Zenodo: MARCXML, Dublin Core (according to OpenAIRE Guidelines), DataCite, DCAT, JSON-LD (Schema.org).

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?

Any project vocabularies or ontologies will be detailed in deliverable D4.1 (Field work protocols for in situ data collection and review of accuracy assessment methods). These will be publicly available to allow potential re-use in other projects. When possible, the ontologies and vocabularies will be linked to commonly used ontologies.

The content of this section will be updated later as D4.1. becomes more complete.

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

A qualified reference is a cross-reference that explains its intent (https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/). Whenever possible, the data generated in the MultiMiner project will include references that increase the interoperability of the data. Such references include, but are not limited to, instrumentation and other datasets produced by the project.

3.3 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.)?

Data reuse will be ensured through careful and extensive documentation in different project phases. The harmonized metadata for *data acquisition* are detailed in D4.1. During the *data analysis* phase, the steps taken to process the data will be described in the metadata fields of the data repository and when needed, in additional readme-files.

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?

Yes, the MultiMiner project aims at widest possible re-use of the data by open and immediate sharing of the research outputs and through the use of permissive licenses (see sections 3.2.1-3.2.2 for details).

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

The MultiMiner project aims to make the project data as re-usable as possible through careful documentation, permissive licensing and immediate publishing of the research outputs. Thus, the research outputs of the project are available to third parties during and after the project.

3.3 Increase data re-use

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

Provenance is defined by W3C as "information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability, or trustworthiness" ([link](#)). In the context of RDM, provenance involves detailed descriptions on elements like authors, data acquisition and processing protocols.

In MultiMiner this information will be documented in the metadata fields of the data repositories of choice and when needed, in additional README-files that will be distributed with the research outputs.

Describe all relevant data quality assurance processes.

Any relevant quality assurance practices will be described in deliverable D4.1.

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.

Other research outputs (than data): software, algorithms and models are considered in the relevant sections above.

The allocation of resources is discussed in Section 3.5 and data security is discussed in section 3.6.

3.4 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.).

The MultiMiner project will, by default, share all of its **digital research outputs**, including, but not limited to, data software, codes and models.

The MultiMiner project will also share its **physical research outputs**, i.e. the samples. The specific rules related to this are:
- The samples will only include those that have been collected in the MultiMiner project and for which there are associated research outputs (data, software, codes or models).
- Any additional samples cannot be collected by the general public from the privately owned mine sites of the project.
- Samples that perish (e.g. pieces of vegetation, water) cannot be shared.
- Samples will only be shared for as long as re-using them does not destroy them (unless a permission to destroy them is granted).
- Those interested in re-using the samples will, by default, collect them from the organization where they are stored at the time of request. The parties requesting the samples will pick the samples up and return them to the said organization, unless otherwise agreed.

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.

3.4 Other research outputs

Other research outputs (than data) are discussed wherever relevant in the sections above. Research ethics is discussed in section 3.7.

3.5 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.) ?

The Consortium Partners are dedicated to using approximately 5% of their working time for RDM. This work is done in the context of individual tasks and hence, is covered by the overall project budget. The Consortium Partners will make sure tasks / activities related to data management (downloading, preprocessing, databases etc...) is in the order of magnitude of 5% of the total working time and will organize work accordingly. The project will be using the free data repositories Zenodo and GitHub for sharing its data and software/algorithms, respectively. These data repositories will also act as the long-term data preservation solution of the project. Furthermore, data storing will be handled using institutional cloud storage solutions and there will not be any direct costs to the project.

Data security is discussed in Section 3.6.

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)

There are no direct costs related to making the data and other research outputs FAIR. See the previous response.

Who will be responsible for data management in your project?

Following good RDM practices is everyone's responsibility in MultiMiner. The detailed distribution of work is described in Section 2.4.

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)?

By default, all research outputs of the project will be kept for as long as they have scientific value and a minimum of 10 years. Data, which are deemed not to have specific scientific value, for instance spurious data, will be discarded at the end of the project. By default, each consortium member will evaluate the usefulness of the datasets they themselves have acquired and will follow the data policies of their respective organizations in terms of data storage times.

The long-term preservation of all the data of scientific value will be secured by sharing them in Zenodo. Zenodo guarantees that data and metadata are retained "for the lifetime of the repository", which follows the lifetime of "the host laboratory CERN…defined for 20 years at least". GitHub also preserves its data and codes in its Archive Program, which aims at long-term preservation of the individual datasets. The GitHub archive program furthermore stores "multiple copies across different data formats and locations" (https://docs.github.com/en/repositories/archiving-a-github-repository/about-archiving-content-and-data-on-github).

3.6 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)?
The MultiMiner project does not have any sensitive data in the sense of GDPR (link). It does have confidential data as listed in Table 6. All confidential and non-confidential data will be stored in the secure servers of the project. Data that are perceived to have potential scientific value, will be kept for a minimum of 10 years after the project. The secure servers have back-up mechanisms to prevent accidental loss of data. Any data transfers of confidential data will be done using secure connections (e.g. encrypted and decrypted using private keys).
Will the data be safely stored in trusted repositories for long term preservation and curation?
Yes, see section 3.2.2. By default, all data and other research outputs will be shared in trusted repositories and data that will *not* be shared (e.g. confidential data) will be kept in the secure servers for a minimum of 10 years or for as long as the data has scientific value.

3.7 Ethical, legal and societal aspects

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 MultiMiner project has no specific ethical issues since the project does not include research involving humans, animals, or other questions that would require reviews by ethics committees. In general, the project is committed to following the European Code of Conduct for Research Integrity (https://allea.org/code-of-conduct/).
It is possible that during the project, stakeholder surveys will be carried out. The surveys will be conducted anonymously, meaning, that they do not have any personal data in them (no open-ended questions or questions that alone or together might allow for any individuals to be identified).
Personal data (name, email address) may be collected in the context of registrations to public events and/or newsletter subscriptions. Here, care will be taken that those whose personal data are collected are informed about the processing of their personal data e.g. through Privacy Notices and that any applicable laws e.g. the General Data Protection Regulation (GDPR) will be strictly adhered to. Any personal data will only be stored for as long as is strictly necessary, for instance, until the end of an event (in the case of event registrations) or until the end of the project (in the case of the project newsletters).
Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?
The MultiMiner project will not have questionnaires that would collect personal data.

3.8 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 MultiMiner project will follow institutional, national and international data policies, recommendations and guidelines at multiple levels: • **Institutional**: any relevant *data/open science policies* of the partner organizations will be followed. It is *required* that each partner ensures that there are no conflicts between this document and any institutional policies at any time.

- **National**: any relevant *national guidelines* or *recommendations* of the partner organizations will be taken into consideration as far as it is feasible and relevant. It is *recommended* that the Consortium Partners will follow the open science agenda of their countries and will bring any potential changes for discussion with the other partners.
- **International declarations and recommendations**: the project will take into consideration any relevant international recommendations and declarations, for instance:
    - UNESCO Recommendation on Open Science
    - Declaration of the 1999 UNESCO World Conference on Science
- **Legislation**: If in doubt, the Consortium Partners will consult the legal departments of their institutions in any matters related to data, patents, IP rights or other matters that may arise with respect to their data and other research outputs.

The open science and RDM guidelines of the European Commission will be strictly adhered to during the project.

4. The Research Outputs of the Project

In this section, all research outputs of the MultiMiner project will be collected. Research outputs include, but are not limited to, publications, data, software, codes and models. **This section will be updated as the project progresses**.

4.1 Data

To be updated as the project progresses.

4.2 Software and algorithms

To be updated as the project progresses.

Table 6 The MultiMiner datasets, their origin, formats, sizes and level of confidentiality. TBD=to be defined.

	Dataset/type of data	Origin	Format	Size	Confidential	Rationale for confidentiality	Comments
1	Point spectral data	acquired	ASCII	TBD	no	N/A	
2	Proximal / laboratory hyperspectral imaging data	acquired	image	TBD	no	N/A	
3	Ground-Based Radar (GBR)	re-used	TBD	TBD	no	N/A	
4	LiDAR	acquired	.LAZ		no	N/A	
5	Drone InSAR	acquired		TBD	no	N/A	
6	Drone thermal	acquired	*.jpg	TBD	no	N/A	
7	Drone hyperspectral	acquired	image	TBD	no	N/A	
8	Airborne (manned) hyperspectral	acquired	image	TBD	no	N/A	
9	Drone gamma-ray	acquired	TBD	TBD	no	N/A	
10	Drone magnetic	acquired	TBD		no	N/A	
11	Landsat 5/7/8	re-used	image	TBD	no	N/A	
12	Sentinel-1	re-used	.tiff	TBD	no	N/A	
13	Sentinel-2	re-used	.tiff	TBD	no	N/A	
14	Sentinel-5 / Sentinel-5P	re-used	.tiff	TBD	no	N/A	
15	MODIS	re-used	image	TBD	no	N/A	
16	WorldView-3	re-used	image	TBD	no	N/A	
17	EnMAP	re-used	image	TBD	no	N/A	
18	PRISMA	re-used	image	TBD	no	N/A	

19	EMIT	re-used	TBD	TBD	no	N/A	
20	TerraSAR-X	re-used	TBD	TBD	no	N/A	
21	ALOS Palsar	re-used	TBD	TBD	no	N/A	
22	InSAR S1	re-used	TBD	TBD	no	N/A	
23	InSAR TSX	re-used	TBD	TBD	no	N/A	
24	In-situ data – background data from Siilinjärvi	re-used	Shapefile *.dwg, *.obj, *.dtm, *.str	700 MB	yes*	Protecting trade secrets	Existing GIS data from the Siilinjärvi study site

*) If used in any publications, consent is needed from Yara.