M4MINING

D1.2

Data Management Plan V2

Funded by the European Union

This project has received funding from the European Union's Horizon Europe Research and Innovation Program HORIZON-CL4-2022-RESILIENCE-01 under Grant Agreement No 101091462

Project

Acronym:	**m4mining**
Title:	Multi-scale, Multi-sensor Mapping and dynamic Monitoring for sustainable extraction and safe closure in Mining environments
Coordinator:	NORCE Norwegian Research Centre AS
Reference:	101091462
Type:	Research and Innovation Action (RIA)
Program:	Horizon Europe Framework Programme (HORIZON)
Theme:	HORIZON-CL4-2022-RESILIENCE-01-08 Earth observation technologies for the mining life cycle in support of EU autonomy and transition to a climate-neutral economy
Start:	1 January 2023
Duration:	36 months
Website:	www.m4mining.eu
LinkedIn:	m4mining
Consortium:	**NORCE Norwegian Research Centre AS (NORCE),** Norway **Norsk Elektro Optikk AS (NEO)**, Norway **Helmholtz Zentrum Potsdam Deutsches GeoForschungsZentrum (GFZ)**, Germany **Luleå Tekniska Universitet (LTU)**, Sweden **Prediktera AB (PDK)**, Sweden **Geological Survey Department (GSD) of Ministry of Agriculture, Rural Development and Environment of Cyprus (MOA)**, Republic of Cyprus (MOA-GSD) **Panepistímio Patrón (UoP)**, Greece **University of Queensland (UQ),** Australia **ReSe Applications LLC (ReSe)**, Switzerland

Deliverable

Number:	**D1.2**
Title:	**Data Management Plan V2**
Lead beneficiary:	NORCE
Work package:	WP1
Dissemination level:	Public (PU)
Nature:	Data Management Plan (DMP)
Due date:	30 June 2024
Submission date:	12 June 2024
Authors:	Edmond Hansen, **NORCE** Simon Buckley, **NORCE** Robert Clarke, **NORCE** Agnar Sivertsen, **NORCE** Friederike Körting, **NEO**
Reviewers:	Nikolakopoulos Konstantinos, **UPATRAS**

Acknowledgement: This project has received funding from the European Union's Horizon Europe Programme under Grant Agreement No 101091462.

Table of Contents

1. Executive Summary

Deliverable D1.2 describes the second iteration of the Data Management Plan for the m4mining project.

The m4mining project aims to provide an integrated remote sensing approach for mapping and monitoring active and inactive mining sites. This Data Management Plan includes the requirements for a secure and compliant use of data in the research and administrative activities of the project, as well as outlining a plan for later release of appropriate data under **F**indability **A**ccessibility, **I**nteroperability and **R**euse (FAIR) principles.

In the project, knowledge, hardware and algorithms/software for hyper- and multispectral imaging combined with LiDAR for 3D surface measurement are being developed. The application area of the developed technology is material identification from the detailed mine face to the site scale using unmanned aerial vehicles (UAVs) and satellite sensors. Therefore, as part of the development and testing of the combined technology two main types of data are generated. The project generates primary multi-sensor datasets from hyperspectral, multispectral, LiDAR and associated navigation and situation sensors on board UAV platforms developed by the project. This is supplemented by validation data acquired using conventional UAV or lidar equipment, mineral or geochemical analysis, and ground/laboratory-based spectral measurements of samples. Finally, existing, and freshly acquired data from hyper- and multispectral imaging satellites including EnMAP, Sentinel-2, and WorldView-3 are utilized for transferring and upscaling developed processing and analysis methods between mine face and site scales.

The project's internal data storage utilizes existing in-house storage and backup solutions at each partner's location. Large volumes of data (several terabytes) are acquired during field work and are shared through physical exchange of disks. Data deemed potentially relevant for further research or included in publications will be published via the Zenodo repository, pending necessary permits from involved landowners or mining companies. All satellite data used in this project are sourced from research-accessible websites and are already FAIR.

The Horizon Europe Model Grant Agreement requires that a Data Management Plan ('DMP') is established and regularly updated. This plan has been developed using the published DMP template for Horizon Europe beneficiaries and addresses the requirements for research data management of Horizon Europe as described in Article 17 in the Annotated Grant Agreement.

2. Introduction

The Data Management Plan (DMP) is an important element in m4mining and includes the requirements for a secure and compliant use of data in the research and administrative activities of the project. This document provides an overview of the DMP deliverable, outlining the objectives, principles, and strategies for effective data management throughout the project.

In the context of m4mining, the DMP aims to ensure the integrity, accessibility, and usability of research data related to the development of multi-scale, multi-sensor mapping and dynamic monitoring in mining environments. The project's objectives address technological advancements, but also encompass sustainable extraction practices and safe closure techniques.

The DMP covers essential aspects of data management, including data collection, storage, organization, metadata creation, data sharing, and data preservation. Ethical and legal considerations regarding data protection and intellectual property rights are also addressed to ensure compliance.

Project participants have defined roles and responsibilities outlined in the DMP, emphasizing the importance of data documentation, versioning, and quality assurance. The plan also highlights the significance of data sharing, with mechanisms for data discovery, access, and appropriate protection and licensing.

Furthermore, the DMP outlines procedures for data preservation, encompassing long-term storage, archiving, and the selection of suitable repositories. The goal is to ensure the enduring value and accessibility of m4mining data for future research, innovation, and advancements in multi-scale mapping, dynamic monitoring, and sustainable mining practices.

In summary, the DMP for the m4mining project establishes guidelines and strategies for effective data management. By adhering to this plan, the project aims to maximize the impact of its data, supporting multi-scale mapping, dynamic monitoring, sustainable extraction practices, and safe closure techniques in mining environments.

3. Data Summary

In general, the m4mining project encompasses two main types of data: 1) primary data generated by the project team itself (from UAV acquisitions with supporting ground truth/validation data), and 2) reuse of existing datasets (mostly related to satellite imagery). A third category relates to administrative and associated personal data required in project management and contact with the industry through the Minerals Industry Advisory Board (MIAB).

The tables below present a data summary detailing the research needs of data collection, generation and/or processing. For each WP, the dedicated table highlights the material objective of the research activities performed, the type(s) of data managed, the origin(s) of this data, and the purpose(s) of the processing.

WP1 – Project Management and coordination	
Objectives	Efficient technical, administrative, and financial management of the project by monitoring progress in the individual work packages, respecting timeline and budget as well as punctual and appropriate reporting to the European Commission.
Data	Personal data (e.g. personal information, contact information, business personal information of the partners' team members of the participants' research activities and relevant information related to the MIAB in WP9).
Origins	Primary data collection from partners (contacts and business personal information of the researchers involved in m4mining).
Purposes	To establish a platform that will facilitate the efficient communication between partners and be the repository for the project documents and data. To ensure the ethical and legal compliance with the applicable principles, laws, and standards. To enable reporting of the project over its duration.

WP2 – System specification	
Objectives	Establish the functional and performance requirements for the drone system, and the safety regulations applicable at the study sites.
Data	No data expected. This WP outputs are summarised in reports/white papers.
Origins	N/A
Purposes	N/A

WP3 – Optimised UAS platform for mining	
Objectives	Develop an integrated solution for hyperspectral data and 3D surface data acquisition from oblique UAS and ground-based platforms, within the safety and operational requirements for mining sites. Develop algorithms and protocols for safe airborne deployment of the full UAS solution in mining sites.
Data	Non-personal data (e.g., operations contextual data), in the form of: Raw sensor data acquired in lab and field testing, including • hyperspectral sensor data • navigational data, • LiDAR data Postprocessed raw sensor data, including • georeferenced hyperspectral reflectance data • georeferenced 3D surface models
Origins	Primary data collection by the partners and researchers from instrument testing in lab and field.
Purposes	The purpose of this data collection is analysis of the instrument performance on a module and system level as part of the development and testing of the software and optimisation of the hardware configuration.

WP4 – Real-time processing	
Objectives	A streamlined processing system for fast and efficient geometric processing and atmospheric correction of hyperspectral data is to be created based on the specification defined in WP 2. The real-time processing is defined such that the end users are able to receive initial, targeted results directly during data acquisition. Furthermore, a high precision workflow will allow to produce completely processed results on the day of data acquisition
Data	Hyperspectral point cloud and surface model/visualisation data (non-personal data), with format defined as part of the WP (T4.1). Anticipated 5-20 TB (depending on size of test sites, to be defined in WP6). Processing in WP4 may (at least) double the amount of data.
Origins	Raw data to be provided by other partners in the consortium through WP6, processed data generated with PARGE and DROACOR (ReSe Applications LLC software)
Purposes	Validation of same-day processing system, validation of real-time processing capabilities. This is sample data for development, testing and proof of concept only.

WP5 – Visualization Software	
Objectives	WP5 is the critical link between the developed UAV–satellite analysis and stakeholders in the proposed application chain, allowing just-in-time decision making tools and presentation of results to users in the mining life cycle. The system is based on capabilities to be developed and added to NLIVE. NLIVE is used as the data backend, and for presentation of integrated project results (UAV & satellite data, regional and complementary field data) with less critical response times for site management and monitoring). A front end developed and built into NLIVE will be used for the interactive visualization of dynamic 3D mine face data (surface models with thematic spectral analysis results) streamed during UAV surveys for on-site geoscientists and surveyors to gain spatial insights into mineral distribution and content, as well as redirect the target area during flight to ensure correct coverage and data quality.
Data	3D mesh/point cloud data and thematic image results (e.g. surface slope, spectral data quality, material mapping results) with format defined in WP4 and acquired as part of WP6 (non-personal data). Supplementary data from several sources will be combined for visualisation of case study results: • Ground truth/validation and other field data (WP6) - primary

	• Publicly available GIS data (maps, aerial photos etc) - secondary • Data from site owner/operator (e.g. geological maps, field analysis or other proprietary data) - secondary • Satellite imagery from WP6 - secondary
Origins	Primary data from the consortium through WP6 (UAV data acquired in real-time or near real-time; ground truth/validation; supporting geological or environmental data). Primary data from operators/site owners (geological, environmental, safety etc). Secondary data in the form of geospatial (georeferenced) imagery or vector data for visualisation purposes. Examples are topographic maps, aerial imagery, or thematic maps available from public bodies. In addition, satellite data from WP6 will be a source of secondary data.
Purposes	Development of capabilities for visualisation of real-time data acquisition workflows. Development of user interface for supporting high quality UAV data acquisition. Supporting case studies through visualisation of WP outputs together with supplementary site data for presentation to stakeholders, dissemination of results, and for other communication and exploitation activities. The purpose of the data is hence to aid in the development, testing, validation, and demonstration of the technology.

WP6 – Data collection and processing	
Objectives	Acquire multi-source, multi-sensor data from satellite, UAV, and field and laboratory instruments from each of the nominated case study sites. Ensure that the collected data meets the quality requirements necessary for processing and analysis.
Data	• EnMAP hyperspectral data is available at Level 1B (L1B), Level 1C (L1C), and Level 2A (L2A). L2A is the surface reflectance data needed for the m4mining studies. Data is available from EOWEB® GeoPortal in GeoTiff, BSQ & metadata format. BSQ format will be used as it is easily readable by EnMAP Box and ENVI software. The size of a single EnMAP scene in L2A format can be around 650 MB. The size of the compressed RAR file delivered to the user varies between 350-500 Mb. • SENTINEL-2 multispectral data is available to users in SENTINEL-SAFE format, containing image data in JPEG2000 format, auxiliary data, and metadata to read the file in the SNAP software. 'SAFE' is an ESA format for archiving earth observation data. A SENTINEL-2 product folder contains a metadata file in XML format that describes the product's organization and content. The GRANULE subfolder contains image data in JPEG2000 format. The AUX_DATA subfolder contains auxiliary files embedded in the product. The size of a single Sentinel-2 scene in L1C format (compressed) can be between 750-800 Mb. The data of this instrument are acquired on a regular basis at 5 days interval and is useful to a range of users interested in geology. • WordView-2 or 3 and similar, non-public satellite data for the case study sites will be requested from the ESA as part of a third-party mission (TPM) data request for archive or newly tasked data. It is impossible to say now of data access will be granted and data will be available for m4mining. • Hyperspectral UAV data acquired with the m4mining Mjolnir camera system. The system simultaneously acquires Lidar, VNIR and SWIR hyperspectral imaging data and navigation data from the integrated navigation system. Hyperspectral raw data is available for VNIR and SWIR separately and will further be processed in WP4 (atmospheric correction), WP5 (data visualisation software) and WP7 (data analysis). The size of the data depends on the length of the acquired flight line and ranges between 2-8GB (SWIR) and 10-20GB (VNIR) for the raw data. The data consists of a .hyspex file and an associated header file (_raw.hyspex and _raw.hdr). A log file (txt format) is written alongside the data acquisition, compiling meta data of the flight datasets.Navigation data is acquired in T04 or similar data format by the navigation system (T04 rover and base data). Lidar data is acquired per flight line and is >1GB per line, available in. pcap data format. • Hyperspectral imaging VNIR and SWIR laboratory data of representative samples from the case study site. Data acquisition in lab conditions results in HSI data cubes for the VNIR and SWIR camera separately, comprising one raw. hyspex and associated .hdr

	file. SWIR raw data files are commonly <1GB per scan, VNIR raw data is commonly below <8GB per scan. The raw data is corrected to reflectance (.img + .hdr) with an in-house routine by HySpex, resulting in <1GB SWIR .img data and <16GB VNIR .img data files per scan. For a sample set <100 samples storing both the raw and ref data for both the SWIR and VNIR results in total file sizes of around 200GB per sample set. Hyperspectral in-situ point spectrometry, collected on site, representing spectral information of a point source (usually referred to as "ground truthing"). Collected alongside with the GNSS position of the sampling point and can include and RGB image of the sampled surface. • Data associated with case study sites: include mineralogical, textural, and chemical data. These datasets are collected within WP6 by the consortium or existing data from previous projects are being requested and checked (QAQC) for their further use within m4mining for training and validation purposes.
Origins	Primary data, such as UAV HSI data, field spectral sampling and sampling for laboratory analyses will be collected by the partners at the case study sites and in the laboratory facilities. Secondary, public satellite data can be directly requested from the EOWEB® GeoPortal (https://eoweb.dlr.de/egp/). Non-public satellite data will be inquired about as part of the ESA TPM access (https://earth.esa.int/eogateway/catalog/worldview-3-full-archive-and-tasking). Secondary existing data available has been collected from ground sampling campaigns, as part of ongoing research engagements by the partners and associated mining companies.
Purposes	Satellite data such as EnMap and Sentinel-2 data acquired from orbit is intended to cover the test sites at medium spatial resolution enabling the study of the targets in deposit to regional scales and enabling a temporal analysis of the targets over time. The UAV HSI data and associated data will be used to survey the case study sites. It will be calibrated, processed, and analysed (WP4, WP7) to derive interpretations of mineralogy and ground chemistry. The existing mineralogical, textural, and chemical data will be used to validate interpretations from the HSI data (WP8). Both the existing data and the primary data collected on site within m4mining (UAV HSI data, geological data, samples) will be used to test and develop potential analysis workflows (AI, Machine learning, and other) for the creation of high-resolution material maps and to upscale and improve satellite data sets.

WP7 – Data analysis	
Objectives	Implement and apply a suite of spectral processing algorithms transferable between multiscale and multi-resolution spectral data, providing key spectroscopic products such as: mineralogic (and lithologic) classification maps, composition and abundance maps (inc. ore grade estimation where applicable); crystallinity maps for mineralized and non-mineralized zones and/or altered zones and host rocks in a quantitative to semi-quantitative fashion; spectral feature characteristics of pH-sensitive iron-oxides and hydroxides aiming to map and track pH environments and acid mine drainage.
Data	Hyperspectral and multispectral imaging data collected from UAV and satellite (.img, .geotiff, …) Hyperspectral point spectrometer data (.asd, .csv, .txt) • Laboratory-based hyperspectral data of samples from WP6 (.img) • Archive spectral mineral libraries (.txt, .csv) • Data relating to the ground truth (GNSS position of sampling points, existing data from previous data generation outside of m4mining) The data formats will be handled as tables (for example .txt, . csv), and imaging files (for example .raw or .img) within WP7
Origins	New data (primary) collected during field work in m4mining (WP6), and existing data (secondary) available from various sources that is re-used by the partners in m4mining. Hyperspectral VNIR and SWIR UAV data collected at case study sites in WP6, processed to at-surface-reflectance in WP4, including Lidar data (DSM) acquired alongside the hyperspectral data, primary data collected by consortium partners at the sites in WP6. Both UAV and laboratory-based hyperspectral data is addressed. Primary hyperspectral point spectrometry, acquired as ground truth and likely used within a spectral library file (.txt or .csv or .asd format) alongside GNSS position data (GPX or GPS exchange format or similar, depending on the handheld GNSS device). Secondary data available and associated with the UAV and satellite data. This data is linked to the UAV and satellite-based data either via GNSS position or from sampling points and associated laboratory analyses. Primary data of the UAV campaigns including data for communication and dissemination purposes like videos, photos, and other means of documenting the activity of the operation.
Purposes	The data will be used for developing and testing different data analysis methods, algorithms, models, and software. The site-specific material and surface maps (e.g. mineral maps) created in this WP also demonstrates the full capability of the developed technology, from data acquisition, via real-time and near-real-time processing, to mineral maps provided to the end-user.

WP8 – Data product interpretation and validation	
Objectives	Interpretation of the analysis products in their geological, operational context for the different case studies. Validation of the data products against the ground control points collected in WP6 for both the UAV and satellite scale. Integration of the different hyperspectral and multispectral imagery with each other as well as with relevant mine geospatial, field and subsurface data e.g. boreholes, stability, etc.
Data	Mapping data products derived in WP6 and WP7 (material maps, digital surface models format: geotiff format, las, img, imaging formats like png, jpeg) and sample data from physical field samples (spectral, geochemical, mineralogical – gps, csv, txt, speclib format) Existing data sets from prior sampling campaigns by partners or associated companies and published maps and industry reports e.g. geochemical or mineralogical data (csv, txt, speclib, geotiff formats)) Secondary, non-public data, from the test site and associated companies (pdf, txt, csv, xscl, jpg and other image formats)
Origins	Primary data products derived in WP7 from hyperspectral UAV and laboratory scanning and analysis in WP6. Primary satellite data products derived by the consortium (WP6), based on secondary satellite dataset acquired by different sensors (EnMap, Sentinel-2, WorldView where accessible and other). Secondary data from existing data sets (prior sampling campaigns by partners or associated companies with the case study sites) and published maps and industry reports Secondary, non-public data, from the test site and associated companies
Purposes	Validation of data products for satellite and UAV-based surveys via ground truthing by different formats, including but not limited to: • Geochemical, mineralogical data from existing or m4mining-related sampling campaigns • Site operators and responsible people privy to site-sensitive information • Published reports and data (industry and scientific) where existing and accessible The purpose of this exercise is to compare the derived data products from WP7 with data from commonly used methods (ground truthing data from the site, e.g. geochemistry, pH,) to provide a level of accuracy for the derived data products and maps generated in WP7 and information about the level of validation possible for the used method applied in WP4 and WP7.

WP9 – Industry Impact	
Objectives	Determining the industry impact of the developed system and data products of the UAV and satellite monitoring approaches that are being developed within the project. The objective of this work package is to draw a clearer line between state-of-the-art methodology used within mining right now and the expected improvements that can be achieved by implementing the envisioned solution within these flow sheets.
Data	At this stage it is unclear if there will be any data associated with this Work Package. There will be discussions with industry representatives on the Minerals Industry Advisory Board. There is potential for contextual data, such as descriptions of drone-based operations. Personnel data (non-sensitive) may be collected in relation to the MIAB for efficient administration (see WP1)
Origins	If any data is captured it will be sourced from: • Grey literature (industry reports, industry conference proceedings) • Ad hoc conversations • Published, open-source information (scientific journals or reports)
Purposes	Where non-personnel data is collected, it will be for comparison purposes to provide a baseline against which to benchmark the performance of the developed system. Personnel (non-sensitive) data related to the MIAB will be for the purposes of efficient and effective administration.

WP10 – Dissemination and communication	
Objectives	Maximise the project outcomes and impact through effective exploitation, communication, and dissemination
Data	Photos, videos, interviews (verbal and transcripts), images, figures, and other forms of communication related data to the project in different common formats (mp4, mp3, mov, jpg, png, txt, svg, pdf). Data collected for communication efforts are mainly anonymized statistics collected via Wordpress (WP Statistics plugin) and via LinkedIn. Both are compliant with GDPR. WP Statistics collects information like site visits, visitors, page visit counts, referrals per search engine. User IP is anonymized and hashed; information is not stored. The plugin does not collect or store any data about the user's visits to the website. Reporting about the key statistics is enabled via email to info@m4mining.eu and via the visual dashboard within Wordpress. GeoIP is collected per user, assigning a country IP of 000 to private IP addresses. All data collected within the WP plugin will be purged after 1300 days.
Origins	The data will most likely be primary data collected by the consortium to either communicate results or project-objectives or to disseminate results. This is likely material created in WP8 and WP9 as well as communicating project innovation from WP2, 3, 4 and 5 and showcasing project actions from WP6, e.g., videos and photos documenting the field campaigns and data acquisition campaigns. The data will be of different visual and audio formats and can consists of, but are not limited to, videos, audio collection, info graphics and written content. Website statistics are collected via WP statistics collected for Wordpress are stored locally on the server and anonymized.
Purposes	The purpose of this data is to communicate and disseminate project results and to effectively communicate information to different stakeholders and via different platforms. It ensures a clear communication of the project. Following up with statistics liker follower or website visit numbers and regions enables the consortium to see the impact of their communication and dissemination efforts and to adjust communication accordingly to target people and organisations of interest to the project.

4. FAIR data

Open access to scientific publications: Each beneficiary ensures open access (free of charge) to all peer-reviewed scientific publications relating to project results (in accordance with the terms of the Grant Agreement). In particular, m4mining project participants are committed to Open Access Publishing, and will prioritise publication venues and promote Open Access to its publications.

Open access to research data: Regarding digital research data, the beneficiaries will deposit the data (including associated metadata) needed to validate the results presented in scientific publications in a research data repository, and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the data as soon as possible. As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data if the achievement of the action's main objective would be jeopardised by making those specific parts of the research data openly accessible. Similarly, open access to research data from specific mine sites is pending on the necessary permits from involved landowners or mining companies.

4.1. Making data findable, including provisions for metadata

To facilitate the findability of data and ensure the provision of metadata, the project aims to deposit research data products needed to validate the results presented in scientific publications, such as datasets, publications, and software, in the Zenodo repository, which offers persistent identifiers for long-term accessibility.

EnMAP and SENTINEL data (WP6) will be identified by a persistent identifier. The datasets are created following the in-house standards of the DLR and ESA space agencies which provides rich metadata for discovery. The datasets are searchable in the internal archives based on acquisition dates and geolocation. All satellite data used in this project are sourced from research-accessible websites and are already FAIR.

Hyperspectral imaging data (HSI) sampled by drones and associated outputs (WP6, 7, 8) is stored with its associated metadata (in the form of header or log files). The data is geo-referenced, and each dataset is labelled according to the date and time of the data acquisition, making the data findable. Data will be stored in project- and case study related folder structures. Keywords in the metadata, including standardized file name extensions or similar, and a findable metadata structure to optimize the possibility for discovery not determined at this point will be discussed.

4.2. Making data Accessible

The M4Mining project aims to make the raw research data accessible within the consortium through efficient internal data management practices. To achieve this, the consortium have explored assorted options. Cloud storage and collaboration platforms such as Google Drive and Dropbox were considered for secure and easy data sharing among consortium members. The goal was to identify a low-cost, secure, and user-friendly solution that meets the unique needs of the m4mining project, ensuring efficient internal data management and making the raw research data easily accessible to consortium members. During each field campaign many terabytes of raw data are sampled. The processing may (at least) double the amount of data. Storing and sharing such data amounts by cloud storage platforms are therefore considered prohibitively expensive and not appropriate for the specific needs of the project. The raw data is stored by NEO at their own internal servers following each field campaign. The collected data is distributed between the partners working on the data by exchange of portable drives. The partners involved in the subsequent processing steps store the processed version of the data in their local infrastructure. To ensure open access to research data,Zenodo's ability to securely store and assign DOIs to the data is considered appropriate. The amount of data likely to be made available is still an open question.

WP4 highlights that campaign and hyperspectral metadata is stored with data acquisitions in agreement with the upcoming standard IEEE P4001 (TBD). Keywords in the metadata to optimize the possibility for discovery for potential re-use will not be provided in WP4.

WP6 satellite data is already available to the public via trusted repositories and users can access it through designated data portals or archives. Both the EnMAP and Sentinel-2 datasets are free to users worldwide. EnMAP data is made available to users for scientific, educational, and non-commercial purposes but the redistribution of the data is subject to restrictions. Users may need to obtain explicit permission from the data providers before redistributing the data to third parties. Users are allowed to access, use, and redistribute the Sentinel-2 data to third parties. All m4mining project members have access to the (EnMAP) data with no restrictions but they might not be allowed to share the data to the public. Given the specification of the data (date and location), potential users can access the data using the designated data portals. Identity of persons accessing the data is ensured by the providers by signing up completing some registration details while accessing the designated data portals. This data does not require evaluation by a data access committee.

WP6 case study data involve existing data collected during projects not related to m4mining, with commercial and government partners. Permission will have to be sought to release this data to the public. Likewise, permission to release case study data acquired during the project must be sought from the land owners or mining companies operating at each site. It is intended the data will become available through publications and government reports, that will undergo a QAQC process then be released through Open Data Portal (under the guidance of the Queensland Department of Resources; https://geoscience.data.qld.gov.au/).

As far as possible, the UAV HSI and associated data collected during the m4mining project including lab-based sample data (for example spectral and geochemical analysis) which are needed to validate the results presented in scientific publications, shall be made FAIR through the designated European portals and Australian national portals. It will be available upon request. However, because there will be PhD students using this data, the data will be under embargo for a period.

4.3. Making data interoperable

Use of standard file formats will be adopted where applicable. Metadata as described will be defined to accurately describe data characteristics and provenance, stored per data object to enable reuse and interoperability.

EnMAP and Sentinel-2 data are publicly available from the EOWEB®GeoPortal and considerations for mapping specific ontologies or vocabularies do not apply. The data are understandable to users. The aim is to publish the results obtained from these datasets. The data will not include qualified references to other data.

4.4. Increasing data re-use

m4mining plans to use standard open access licences such as Creative Commons attribution for the data and research outputs of the project. In relation to the research outputs and data products from the project, m4mining will determine the appropriate license by carefully considering the desired permissions and restrictions for the data. This will be achieved by reviewing the available CC licenses, such as CC BY, CC BY-SA, CC BY-NC, and CC BY-ND, and understanding the implications of each. Once the license type is selected a license statement that clearly describes the terms and summarizes them in a user-friendly manner will be created. This statement will be prominently applied to the research data outputs, ensuring its visibility in the documentation, metadata, or alongside the data files. The m4mining project will also ensure compatibility of the chosen license with any third-party content included in the data. Where appropriate, supporting documentation, such as a license guide or FAQ, will be prepared to assist users in complying with the chosen license and providing proper attribution. m4mining will deposit the data in recognized repositories

that support the required open data practices, ensuring effective dissemination and long-term accessibility. Consistency will be maintained by applying the chosen license to all versions and updates of the data. By following this process, the m4mining project will establish a clear and appropriate license for its research data outputs, facilitating wider sharing and reuse while promoting transparency, collaboration, and the principles of open science.

The satellite analysis methodology of WP6 will be made available to the public or published facilitating the validation of data analysis and re-use. Sharing of this data does not imply proprietary codes or algorithms used in the analysis will necessarily be published, but all research results and outputs of m4mining are assessed as part of the dissemination and exploitation work of the project. Open access for publications and FAIR access to data the data products of the project is a core principle of the project. The results in WP6, together with analysis methodology could be shared but the redistribution of raw data (which in any case is publicly and freely available from EOWEB®GeoPortal) might be limited as explained above.

For data produced in WP6, while it is possible that results obtained over a certain site are of interest could be useable to a third party, the methodology to process EnMAP hyperspectral data might be more attractive. The provenance of the data will be thoroughly documented using the appropriate standards. In terms of data quality assurance, which is undertaken by the data providers themselves.

HSI UAV-based data in at-surface reflectance will be made available upon request to enable validation and re-use of the data. This will only be the case after the associated students' work is not affected by sharing data publicly and the case study site managers do not deem the data collected over their site as sensitive.

5. Other Research Outputs

In WP4 both PARGE and DROACOR programs are used and enhanced for this project are ReSe Applications LLC proprietary software. Within the project consortium, workflows, user manuals, and other schematics for the project are shared via the NORCE Teams Sharepoint (in the relevant WP folder).

In WP6 The high-level workflows, analysis methodology and the yielded results (material maps, as specified above) will be made available to the public as part of publication efforts by PhD students working in the project in accordance with the principle 'as open as possible, as closed as necessary' to ensure proper management of IPR and exploitation of the results and outputs of the project.

6. Allocation of resources

Costs of making data and outputs FAIR

Costs in relation to making data and outputs 'FAIR' are directly necessary for the implementation of the project as it is described in Annex 1 for the implementation of the m4mining Grant Agreement. Costs incurred making the data outputs FAIR will therefore be reported by the project partners and covered by the Grant. m4mining will also use the available support and resources highlighted at 4.1 to ensure data and outputs are FAIR over the longer term.

Responsibility for data management

Raw data generated in (WP3) as part of the testing and software development is managed and stored on the servers in place at those partners internal to NEO and LTU.

Hyperspectral data collected within the case studies (WP6) will be processed and managed internally within NEO, GFZ, UQ and UoP. Final data outputs are shared with partners for the analysis (WP7, PDK) to generate data products (maps) and to visualize the data (WP5 and WP7, NORCE, PDK, NEO).

Long Term preservation

WP6 satellite data are free of charge to users. Management of this data used in the project falls within WP6 under the responsibility of the WP6 leader. Long term access to the data is assured by the space agencies (DLR and ESA). WP6 UAV-based HSI data and associated data (e.g., lidar-generated DSM, navigation data, etc.), will be stored by NEO on an internal NAS server, with weekly backup offsite. Access is secured by multi-factor authentication (MFA) behind a firewall. Other data is stored in Microsoft Teams (OneDrive/ Sharepoint) following Microsoft Standard Backup and Recovery. Confidential data will be labelled. Long term access is not foreseen for this data type.

Hyperspectral UAV reflectance data outputs that are relevant for subsequent analysis and interpretation for associated partners, PhDs and students and the case study site managers will be stored with long-term preservation in mind on local NEO NAS server and possibly with the responsible partners.

Research data outputs of the project are planned to be preserved over the long term according to the solutions outlined at 4.1.

7. Data Security

In relation to WP6, the data providers secure raw data.

Internally at GFZ, raw data is secured via regular backups on servers. Data collected by NEO is stored on an internal NAS server, with weekly backup offsite. Access is secured by multi-factor authentication (MFA) behind a firewall. A shared repository

8. Ethics

In developing the data management plan for the 'm4mining' project, we acknowledge the legal, ethical, and intellectual property rights (IPR) protection considerations, as well as the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles, particularly in relation to the project's data products. We will ensure compliance with relevant IPR regulations and respect the intellectual property rights of contributors and stakeholders involved in the project. To adhere to the FAIR principles, we will focus on enhancing the findability, accessibility, interoperability, and reusability of the data products generated by the project. This will involve assigning unique identifiers, applying standardized metadata, and providing comprehensive documentation to facilitate the discovery and utilization of the data products. While promoting openness and accessibility, we will also establish appropriate data sharing and access policies to safeguard sensitive or proprietary information. By embracing IPR protection and adhering to the FAIR data principles for the project's data products, we aim to foster responsible data management practices, encourage collaboration, and maximize the value and impact of the 'm4mining' project.

9. Source Material

Relevant authorities

1. 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 (General . (n.d.).

2. Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union. (n.d.).

3. Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act). (n.d.).

4. Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications). (n.d.).

5. EC. (2023). Horizon Europe (HORIZON) Program Guide. Brussels. programme-guide_horizon_v2.0_en.pdf (europa.eu) (last accessed 12.06.2023)

6. WP29. (2014, April 10). Opinion 05/2014 on Anonymisation Techniques (WP216). Brussels. ??