AGEMERA

Critical Raw Materials for
a Resilient Europe

DELIVERABLE 4.2

Workflow Process for Combining Datasets

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

Project no.	**101058178**
Project acronym:	AGEMERA
Project title:	**Agile Exploration and Geo-Modelling for European Critical Raw Materials**
Call:	HORIZON-CL4-2021-RESILIENCE-01
Start date of project:	01.08.2022
Duration:	**36 months**
Deliverable title:	D4.2 *WORKFLOW PROCESS FOR COMBINING DATASETS*
Due date of deliverable:	30.11.2023
Actual date of submission:	29.11.2023
Deliverable Lead Partner:	**Muon Solutions Oy (MUON)**
Dissemination level:	Public

Author list

Name	**Organisation**
Marko Holma	Muon Solutions Oy (MUON)
Pasi Kuusiniemi	Muon Solutions Oy (MUON)
László Balázs	Muon Solutions Oy (MUON)
Sándor Demők	Muon Solutions Oy (MUON)
Barbara Stimac Tumara	OPT/NET (OPT)
Josipa Kapuralic	University of Zagreb (UZG)
Markku Pirttijärvi	Radai Oy (RAD)
David Marti Linares	LITHICA (LITH)
Giulio Casini	LITHICA (LITH)
Helena Seivane	CSIC (CSIC)

Document History

Version	Date	Note	Revised by
01	2.10.2023	V0.1 Draft	Marko Holma
02	20.10.2023	V0.2 Added descriptions of geological data types	Marko Holma
03	25.10.2023	V0.3 General upgrades by adding text	Barbara Štimac Tumara
04	29.11.2023	V1.0 Final version	Pasi Kuusiniemi

Disclaimer

The content of the publication herein is the sole responsibility of the publishers, and it does not necessarily represent the views expressed by the European Commission or its services.

While AGEMERA is funded by the European Union, views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the European Research Executive Agency (REA) can be held responsible for them.

While the information contained in the documents is believed to be accurate, the authors(s) or any other participant in the AGEMERA consortium make no warranty of any kind with regard to this material including, but not limited to the implied warranties of merchantability and fitness for a particular purpose.

Neither the AGEMERA Consortium nor any of its members, their officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing, neither the AGEMERA Consortium nor any of its members, their officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

Executive Summary

Brief Overview

This report, specifically Deliverable 4.2 (D4.2) of the Horizon Europe AGEMERA project, focuses on elucidating the methodologies and outcomes of Task 4.2 (T4.2), titled “Data Fusion.” The primary objective here is to integrate a variety of datasets from different Work Packages (WPs) to gain an optimal understanding of the geological features of each field trial area. This process involves using satellite-based spatial data as an additional input and builds upon the homogenised data compiled in Task 4.1 (T4.1).

The specific aims of T4.2 include analysing and interpreting individual data layers to enhance the understanding of geological structures and identifying and characterising selected mineral system footprints. This work is essential for a comprehensive understanding of the selected CRM deposits. The report also outlines the data management process, which involves various phases like data collection, homogenisation, integration, fusion, and geological interpretation, using advanced technologies for sustainable resource management.

Objectives and Scope

The report’s primary objective is to thoroughly explain the methodologies and results of T4.2, which is focused on “Data Fusion.” The essence of T4.2 is to integrate various and heterogeneous datasets from the geological and geophysical Work Packages (WPs) 1 and 3, respectively. The aim is to comprehensively understand the geological features within each field trial area. The specific objectives of the report are:

- Analyse and interpret individual data layers to enhance the understanding of geological structures, lithologies, and hydrothermal alteration systematics.
- Identify and characterise mineral system footprints, contributing to a thorough understanding of the selected Critical Raw Materials (CRM) deposits.
- Provide additional tools and methodologies to improve the knowledge of geological, geochemical, and geophysical features of the CRM deposits.
- Establish a linkage between T1.4 and T1.5, which, respectively, are tasks in WP1 focusing on computer-based modelling and developing new and improved mineral system models for CRM ore deposit types.

The scope of this report extends to delivering a comprehensive and integrated workflow on how geological and geophysical datasets have been combined in the AGEMERA project through the application of data fusion methodologies. This workflow is designed to be adaptable, making it applicable for similar data fusion tasks outside the current project’s scope.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

Additionally, the report covers the overall data management process in the AGEMERA project. This process involves multiple stages, starting from data collection and ending with geological interpretation via homogenisation, integration, and fusion. The data collection phase encompasses gathering various types of data such as survey data, laboratory data, geophysical measurements, and satellite data. The homogenisation phase involves transforming the selected data for interpretation and geological modelling, while the integration phase focuses on creating an efficient system for geological interpretation. In the fusion phase, information from different sources is combined into rock-quality information. Particular emphasis is placed on the description of the integration of advanced data types such as drone geophysics, passive seismics, and cosmic-ray muography.

Key Findings

Integrating diverse data types enhances understanding of geological structures and the past mineralisation processes. Advanced methodologies can provide deeper insights into mineral system footprints and establish a vital linkage between computer-based modelling and improving ore deposit models. The used integration methodology is an example how geological research and exploration can be advanced.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

List of Acronyms

ANT	Ambient Noise Tomography
API	Application Programming Interface
CRM	Critical Raw Material
CSIC	Spanish National Research Council
D	Deliverable
DIAS	Data and Information Access Services
GDAL	Geospatial Data Abstraction Library (translator library for raster and vector geospatial data formats)
GIS	Geographic Information System
LITH	Lithica (SME Partner from Spain)
MUON	Muon Solutions Oy (SME Partner from Finland)
NetCDF	Network Common Data Form (a set of software libraries and self-describing, machine-independent data formats)
OCLI	Platform's backend component by OCLI Inc.
OPT	OPT/NET (SME Partner from the Netherlands)
RAD	Radai Oy (SME Partner from Finland)
SME	Small and medium-sized enterprise
T	Task
TT	Tallinn University of Technology (Academic Partner from Estonia)
UNZA	University of Zambia IWRM Centre (Academic Partner from Zambia)
UZG	University of Zagreb (Academic Partner from Croatia)
WP	Work Package

List of Figures

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

1. Introduction

1.1 Background

1.1.1 AGEMERA Project

The AGEMERA project focuses on advancing mineral exploration methods in the European Union (EU) and Zambia to meet the demands of a low-carbon and digital economy. Utilising a combination of conventional and state-of-the-art geological and geophysical surveys, the project aims to map Critical Raw Material (CRM) resources across approximately 4,700 km^2 in seven countries. It will employ three innovative, non-invasive survey techniques — passive seismic methods, multi-sensing drone system, and muon-based multidetector density detection system (utilising a novel astroparticle physics technique called muon imaging or muography) — to enhance the accuracy of existing genetic mineral system models. Concurrently, the project will engage with stakeholders to introduce United Nations Framework Classification (UNFC) guidelines for mineral resources and will undertake extensive public outreach, targeting an audience of 5,000,000 citizens by 2030. Additionally, AGEMERA will develop an open-access SoftGIS database to capture social, cultural, environmental, and economic concerns related to mining, thereby facilitating the creation of socio-economic potential maps to be used alongside geological potential maps. This report focuses on how geospatial data is handled in the project in data fusion of various types of data.

1.1.2 Present Report's Place in the AGEMERA Project

The present report is an output of WP4 of the AGEMERA project. AGEMERA is subdivided into seven (7) Work Packages (WPs): WP1 European CRM Potential, WP2 Impact of CRM on Green & Digital Transition, WP3 Geophysical Data Acquisition, WP4 Data Processing, Fusion & Sharing, WP5 Dissemination, Communication & Collaboration, WP6 Exploitation and WP7 Management. WP4 has solid links to the first three WPs.

WP4 proper is subdivided into five (5) tasks: T4.1 Data processing, **T4.2 Data fusion**, T4.3 Data sharing interfaces, T4.4 Data applications and T4.5 Comparative validation of applications developed by mapping with results of seismic profiling, geological profiles and numerical calculations. The present report is a deliverable of T4.2. This task is led by MUON and involves significant contributions from a consortium of partners RAD, OPT, UZG, UNZA, and TT. However, T4.2 is a collaborative endeavour involving also multiple other consortium partners contributing geological data under WP1 and geophysical data under WP3. Additionally, commercial satellite-based spatial data has been acquired to supplement these datasets.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

1.1.3 Study Areas

The AGEMERA project concentrates its research and technology tests in the following geological areas, illustrated in Figure 1, both as Areas of Interest and the pilot sites.

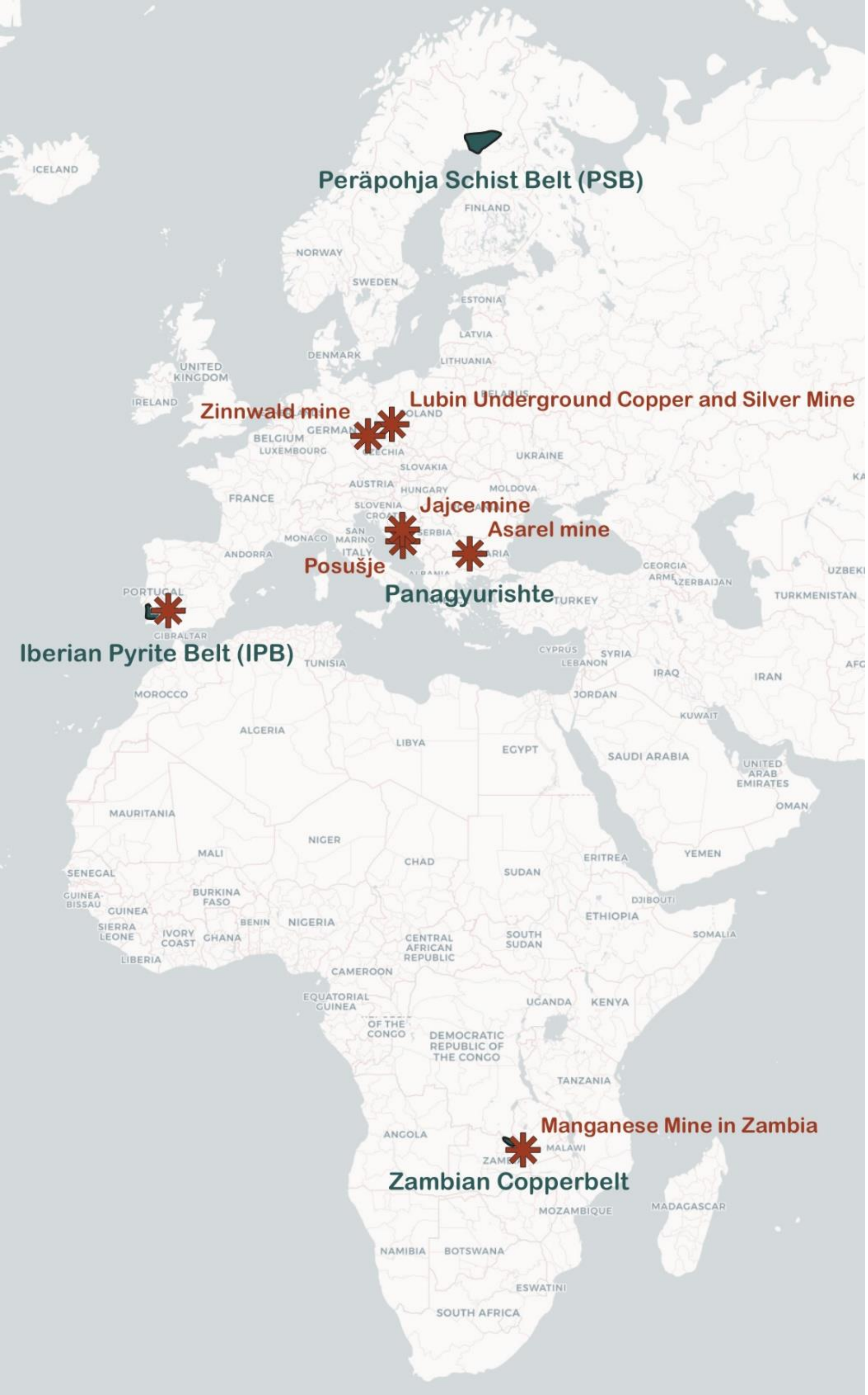


Figure 1. Project's main target regions (polygons) and selected key trial sites (asterisks)

(Reference: Muon Solutions Oy)

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

1.2 Objectives of the Report

The primary objective of this report, which serves as Deliverable 4.2 (D4.2) of AGEMERA, is to elucidate the methodologies and outcomes of Task 4.2 (T4.2), titled “Data Fusion.” T4.2 aims to integrate diverse and heterogeneous datasets from WPs 1-3 to achieve an optimal understanding of the geological features of each field trial area. Satellite-based spatial data is used as an additional input. This integration process builds upon the homogenised data compiled in Task 4.1 (T4.1) (see Section 1.1 for further reference). While T4.1 focuses on the initial combination of some geophysical datasets, T4.2 extends this by including and merging additional geological and geophysical information from WPs 1 and 3, respectively.

The specific objectives of T4.2, and by extension this report, are as follows:

- To analyse and interpret individual data layers to enhance the understanding of geological structures, lithologies, and hydrothermal alteration systematics.
- To identify and characterise mineral system footprints, thereby contributing to a comprehensive understanding of the selected Critical Raw Materials (CRM) deposits.
- To provide additional tools and methodologies that can improve the understanding of geological, geochemical, and geophysical features of the CRM deposits.
- To establish a linkage between T1.4 and T1.5, both of which are tasks in WP1. T1.4 focuses on computer-based modelling, whereas T1.5 is concerned with the development of new and improved mineral system models for CRM ore deposit types.

By fulfilling these objectives, this report aims to offer a comprehensive and integrated workflow on how geological and geophysical datasets have been combined in the AGEMERA project by deploying data fusion methodologies. Furthermore, the workflow presented herein is designed to be adaptable and should prove useful for similar data fusion tasks outside the scope of the current project.

1.3 Structure of the Report

This report is organised into a series of interconnected chapters and subsections, each designed to provide a comprehensive understanding of the workflow process for combining various datasets acquired in the AGEMERA project. The structure is as follows:

- Executive Summary: This preliminary section offers a concise overview of the report's objectives, scope, and key findings. It serves as a quick reference for readers who require a summary of the report's essential elements.
- Chapter 1: Introduction: This chapter provides background information, outlining the objectives and describing the document's overall structure.
- Chapter 2: Methodology: This section delineates the methods employed for data collection, homogenisation, and integration. It serves as a guide to the technical approaches utilised in the report.
- Chapter 3: Data Categories and Dimensions: This chapter is divided into three main subsections — Primary Data Categories, Data Dimensions, and Metadata. It provides a detailed account of the data types considered in the report, their dimensions, and associated metadata.
- Chapter 4: Data Types: This chapter offers an in-depth look at the specific types of data generated in the AGEMERA project, including geological, geochemical, and mineralogical studies, drone geophysics, passive seismic, cosmic-ray muography, and satellite data provided dominantly from the Copernicus Programme (EU's space programme).
- Chapter 5: Conclusions: This final chapter summarises the key findings of the report, discusses their implications for future research, and offers recommendations.
- References: This section lists all the academic publications, articles, and other sources cited throughout the report.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

2. Data management process

2.1 Introduction to the data management process

The project is working towards the declared objectives (as described in Chapter 1) involves extensive, structured and controlled data collection and information gathering to achieve the intended results. This work can be divided in different phases starting from the data collection phase and ending to the geological interpretation phase via homogenisation, integration and fusion.

Data collection phase - In the first phase, the data are collected by the institutions (data providers) designated for each task. These data may include available previous survey data and other geological information, laboratory data based on sampling, results of geophysical measurements and satellite data

The data include details of the measurement and sampling (metadata): coordinates, timestamps, units, instruments and their settings, topographic data, quality and quality control data. The data are processed, classified and checked by experts from the institutions in accordance with the requirements of central integrated storage. At the last stage the data are also stored at the data provider's storage with appropriate logging to ensure traceability.

Homogenisation phase - The selected data necessary for interpretation and geological modelling are transformed (data format, coordinate system) according to the rules of the central data collection (defined by OPT NET) and submitted to the central database for uploading together with detailed descriptive data (metadata).

Integration phase - The verified data system with its established data structure and user interface, provides data for the further efficient geological interpretation. All data required for interpretation are available in a single coordinate system and can be graphically merged. As a result of pre-processing the data system is now free of measurement contingencies and all information can directly be related to rock quality and geological structure. The integrated data system also allows for further verification and easier, clearer archiving.

Fusion phase - The integrated data system allows information from different sources to be combined into rock quality information according to their reliability.

Geological interpretation - The integrated database and the data processing steps that lead to its creation are ultimately aimed at effective geological interpretation. The results of the interpretation can also be loaded up into the integrated database and archived.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

2.2 Data Collection

Data collection in the scope of mining and critical raw material prospecting involves the systematic gathering of geological, geophysical, geochemical and Earth observation data from various sources. This data is essential for understanding subsurface deposits, identifying potential deposits, estimating resource reserves and mitigating negative environmental impacts. Advanced technologies, such as geospatial mapping and machine learning, are increasingly used to streamline data visualisation and analysis for sustainable resource management.

2.2.1 EOD Data collection

In the scope of the AGEMERA project, satellite data is collected via dedicated APIs such as DIAS Finder for Sentinel-1 and Sentinel-2 or Supplier's API for ASTER imagery. Sentinel data (part of Copernicus missions) are collected in SAFE file format (Standard Archive Format for Europe), designed to act as a standard format for archiving and conveying Earth Observation Data within the European Space Agency (ESA) archiving facilities. On the other hand, ASTER data is collected in HDF file format (Hierarchical Data Format), an open-source set of file formats (HDF4, HDF5) designed to store and organize large amounts of data.

The platform's backend component considers the collection and ingestion of different data sources into a data cube representation for a given area of interest (AOI) as an integral step in the workflow.

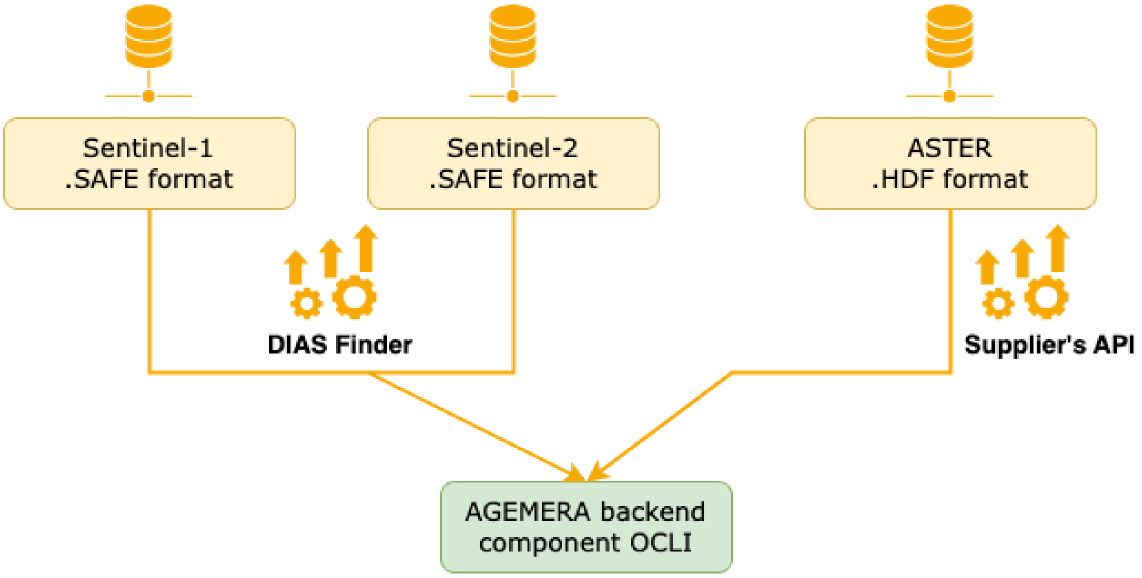


Figure 2. Flowchart for data collection process

Credits: OPT/NET B.V. (2023)

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

2.2.2 2D data collection

The 2D horizontal data processed by the consortium partners (data providers) is collected in GeoTIFF format either as single band rasters or multiband rasters. Multiband rasters need to have an additional fourth band (alpha band or transparency band) incorporated, due to the internal algorithms of the platform's backend component. GeoTIFF is a public domain metadata standard which allows georeferencing information to be embedded within a TIFF file. The potential additional information includes map projection, coordinate systems, ellipsoids, datums, and everything else necessary to establish the exact spatial reference for the file.

All collected data are inspected and stored in the platform's backend component (OCLI) and published with an appropriate AI Knowledge Pack to the platform's frontend component, AGEMERA Geo-Suite, which can then be used as a general data repository for all AGEMERA use cases: Bulgaria, Poland, Spain, Finland, Germany, Bosnia-Herzegovina and Zambia.

2.2.3 3D data collection

The collection of 3D data from the partners who are data providers must be in one of the formats that is routinely used for geoanalytics, appropriate for handling multidimensional rasters, in order to be ingested into the AGEMERA platform and visualised in the AGEMERA Graphical User Interface. Among different file formats that support georeferencing, multidimensional (not only 2D) rasters and are supported by GDAL, the most popular and preferable file format is NetCDF (network Common Data Form) used for storing multidimensional scientific data (variables) displayed through a dimension. It is self-describing, portable, scalable, appendable and shearable data file format. One significant advantage of NetCDF format is that axes (latitude, longitude, height, etc.) may be assigned vectors of coordinates (ticks) that have 1-to-1 correspondence to the coordinates of voxel centres on the actual data grid. This provides an error-free way of establishing the correspondence between spatial coordinates of each cell and its position in the grid of the data array.

Again, as for the 2D data, the 3D files are also inspected and stored in the platform's backend component (OCLI) and published with an appropriate AI Knowledge Pack to the platform's frontend component, AGEMERA Geo-Suite, which can then be used as a general data repository for all AGEMERA use cases: Bulgaria, Poland, Spain, Finland, Germany, Bosnia-Herzegovina and Zambia.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

2.3 Data Homogenisation and Integration

Data homogenization is the process of ensuring that data from different sources or formats are transformed into a common and consistent format for the purpose of data consistency and uniformity. Data integration, similarly, but still distinct from data homogenization, deals more with combining data from multiple sources into a cohesive and uniform dataset. Both processes are crucial for ensuring data quality and enabling effective data analysis. It is important to consider the harmonisation of the products to be developed by the AGEMERA consortium with the INSPIRE directive, most notably interoperability of spatial data sets in the scope of AGEMERA project. Interoperability in INSPIRE means the possibility to combine spatial data from different sources in a consistent manner without involving specific efforts of humans or machines. Interoperability may be achieved by either changing (harmonising) and storing existing data sets or transforming them. Therefore, the aim is to maximise the cross-thematic interoperability of AGEMERA platform outputs utilising INSPIRE spatial data sets and services as well as guaranteeing interoperability with other domains. The Implementing Rules on interoperability of spatial data sets and services (IRs) and Technical Guidelines (Data Specifications) specify common nomenclature, data models, code lists, map layers and additional metadata required for interoperability when exchanging spatial datasets among different processing systems. Several encoding standards are in active use, most importantly GeoTIFF (2D), NetCDF (3D), JSON (1D). In case of AGEMERA platform, it should be distinguished between input file formats (SAFE or HDF for EOD, GeoTIFF for 2D data and NetCDF for 3D data), backends' internal file formats (ENVI, TIFF) and file formats used for storing and visualizing results in Graphical User Interface (COG and GeoJSON), as schematically shown in Figure 3.

Finally, very large sets of information in the AGEMERA Graphical User Interface require better methods for interactions between the computers and human users. Traditional large computer screens, keyboard and mouse interfaces do not provide the easy and natural ways for finding, displaying and analysing very large geospatial datasets. Written and spoken natural language processing (NLP) with AI in English is used with IBM Watson. The support of larger dialog options also makes the interactions with the AI chatbot more natural and human-like.

2.3.1 EOD Data

Earth Observation data acquired either from dedicated APIs Suppliers are ingested into the AGEMERA platform in one of two already mentioned file formats, SAFE or HDF. To be used in further analysis, the file format needs to be adapted and that is achieved with Analysis Ready Data (ARD), a preprocessed assigned time-series stacks of imagery that are produced at a set standard, which allows users to skip straight to analysis. Level 1 ARD (or stack) and level 2 ARD (tensors) are produced internally by the platform's backend in ENVI file format and are used for AI analytics, which produce the result in TIFF file format. After the finished AI analytics, final result is stored in Cloud Optimised GeoTIFF (COG)

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

format in the AWS3 online database for visualisation in the AGEMERA AI Graphical User Interface and as GeoJSON files in PostGIS database with all the accompanying metadata.

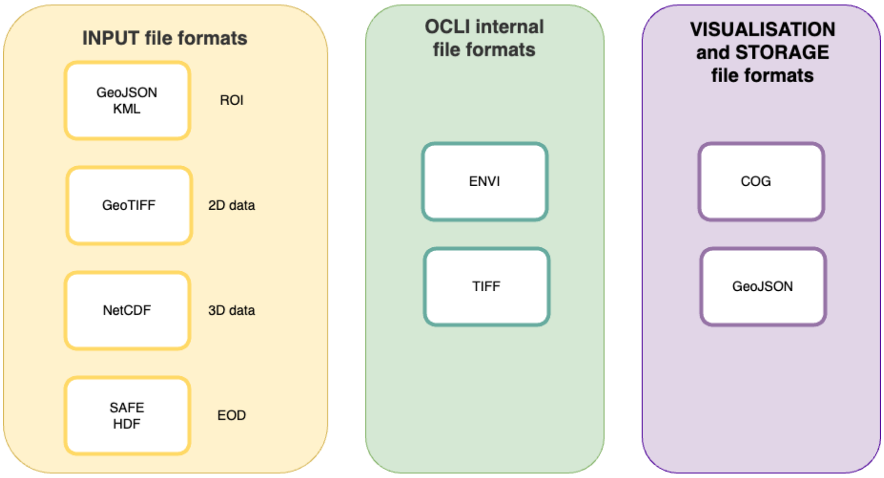


Figure 3. Schema of file formats

(Reference: OPT/NET B.V. (2023))

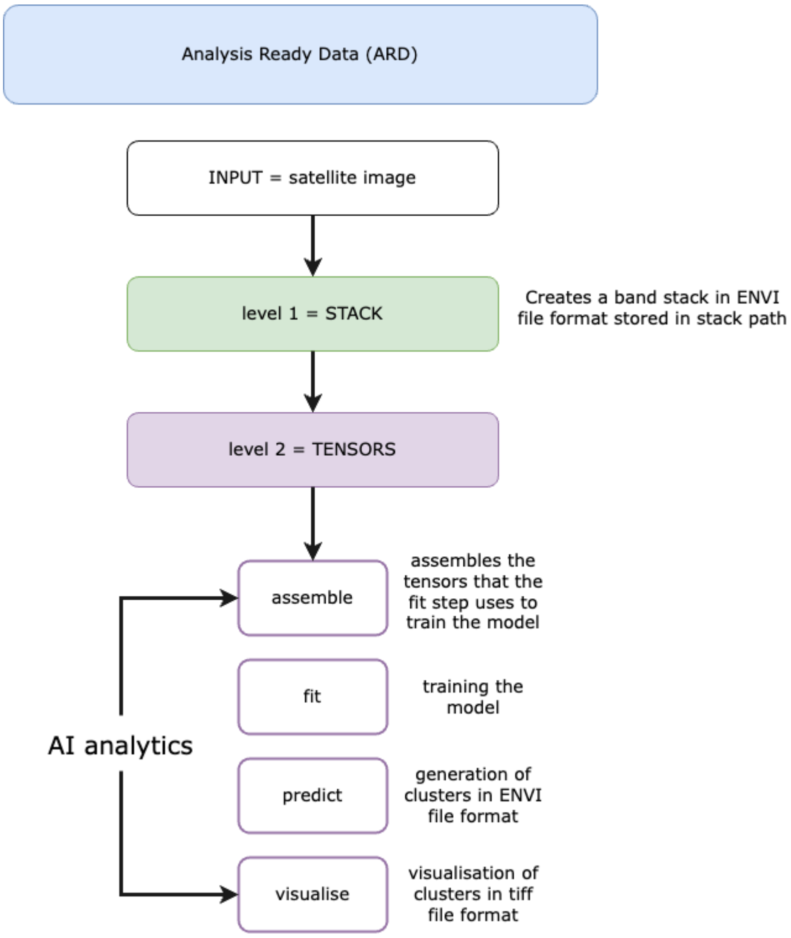

Analysis Ready Data (ARD)
INPUT = satellite image
level 1 = STACK
Creates a band stack in ENVI file format stored in stack path
level 2 = TENSORS
assemble
assembles the tensors that the fit step uses to train the model
fit
training the model
AI analytics
predict
generation of clusters in ENVI file format
visualise
visualisation of clusters in tiff file format

Figure 4. Flowchart of data integration process

(Reference: OPT/NET B.V. (2023))

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

2.3.2 2D data and Coordinate Reference System

As already mentioned, 2D data collected from the partners (data providers) should be in GeoTIFF file format. Considering that, GeoTIFF files shared with OPT/NET need to be properly georeferenced and readable with GDAL. For 2D horizontal data, the preferable is the most basic default standard; WGS84, common latitude-longitude grid, that is standardized as EPSG:4326. Another option is to use the UTM coordinate system which is actually a collection of 60 different specific projections with individual EPSG-codes. One must choose an appropriate one for the location at hand, by converting the UTM zone to corresponding EPSG code. An advantage of UTM is that UTM coordinates are described in meters and roughly correspond to the actual metric distances on Earth surface. This might be useful for performing on-site checks of coordinate correspondence.

Same as for EOD, after the ingestion of GeoTIFF files into the platform backend component with corresponding AI Knowledge Pack either for single band or multiband rasters, the data is stored as Cloud Optimized GeoTIFF (COG) in the AWS3 online database for visualisation in the AGEMERA AI Graphical User Interface and as GeoJSON files in PostGIS database with all the accompanying metadata. For now, there is no possibility of publishing vector file formats, but they can still be temporarily visualised in the GUI with simple drag-n-drop or upload functions.

2.3.3 3D data and Coordinate Reference System

Same as for the 2D data, 3D data need to be properly georeferenced to be ingested in the AGEMERA platform: for NETCDF, it can be provided in metadata according to one of the existing conventions. Unless any additional considerations arise, it might be wise to go with the CF convention that is mentioned as 'the recommended' one.

As for the preferred coordinate reference system, a properly formed NetCDF file must contain georeferencing in such format, that GDAL is able to translate it properly into any other coordinate system with either API or CLI programs. Preferable coordinate reference systems are: EPSG 4979 (ellipsoidal 3D coordinate reference system for 3D objects on the surface of the Earth), EPSG 4978 (geocentric 3D coordinate reference system for the objects above, below and in orbit around the Earth) or EPSG 5773, EPSG 5798 and EPSG 3855, which are vertical datum for EGM96, EGM 84 and EGM2008, respectively. The last three represent only the vertical axis with zero-height surface resulting from the application of the EGM96/EGM84/EGM2008 geoid model to the WGS 84 ellipsoid.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

2.4 Data Integration Process

The aim of the data integration is to collect all information relevant to geological modelling from different sources into a database that gives geological interpretation professionals access and ability to appropriately combine and visualize the information from a map-based database. The data may include the following datasets and structures:

- Map like (2D) datasets
- Sections (2D) datasets (seismic-, resistivity-, density sections)
- 3D (voxelized and categorized) structures

During the integration procedure, remote sensing, surface geophysical, muographic, drilling and other sampling rock data should be placed into the database in a proper structured format to support interpretation. The data management scheme of the institution responsible for providing the (measured or extracted) data in the integration process is illustrated in Figure 5.

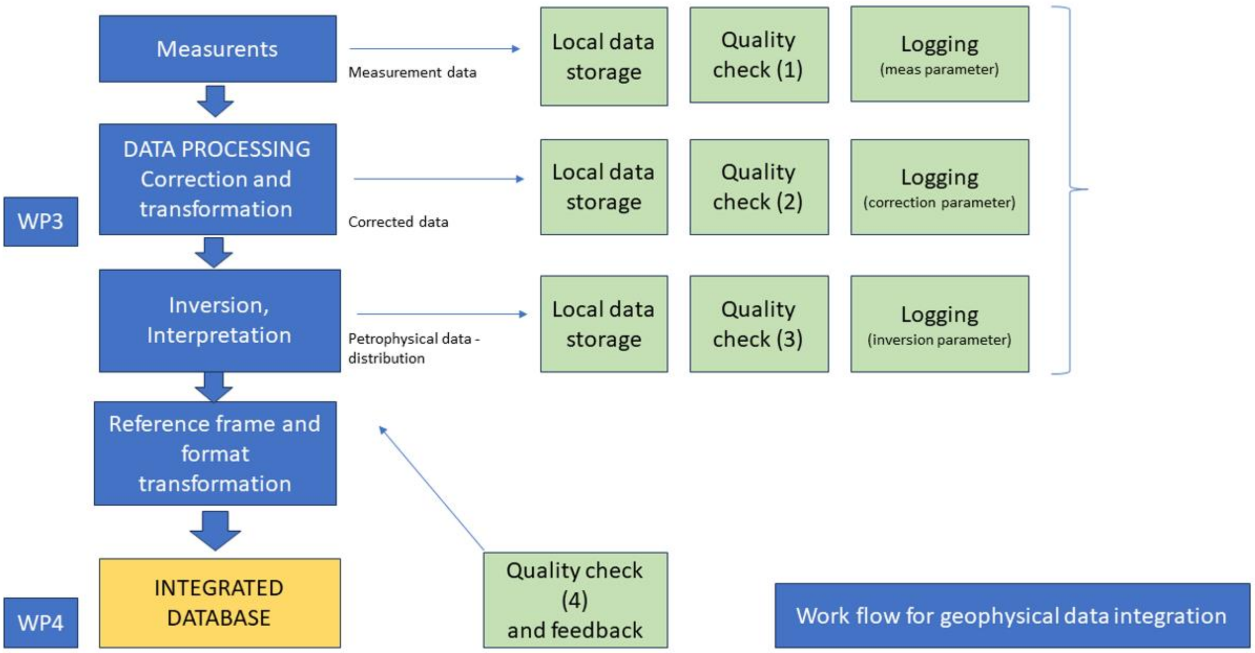

Measurents
Measurement data
Local data storage
Quality check (1)
Logging (meas parameter)
DATA PROCESSING Correction and transformation
WP3
Corrected data
Local data storage
Quality check (2)
Logging (correction parameter)
Inversion, Interpretation
Petrophysical data - distribution
Local data storage
Quality check (3)
Logging (inversion parameter)
Reference frame and format transformation
INTEGRATED DATABASE
WP4
Quality check (4) and feedback
Work flow for geophysical data integration

Figure 5. Data management scheme of the measuring institution in the integration process

(Reference: Muon Solutions Oy)

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

2.4.1 Integration of geophysical data

The geophysical measurement system, defined in the AGEMERA project, consists of different spatial resolution methods with different exploration depths. The results of the field measurements provided by the methods are subject to detailed verification and documented data processing (Castanedo, 2013). The scope of the data integration will include corrected rock physics properties and their spatial distribution derived from the data and directly related to geology (rock quality, texture, structure). These consist of:

- Airborne EM measurement: 2D apparent resistivities (frequency dependent)
- Airborne Radiometry: 2D Gamma intensity map
- Airborne Magnetic measurements: 2D anomaly map (B or B gradient distribution)
- Resistivity tomography: 2D resistivity sections
- Passive seismic: 2D seismic section (on time or depth scale)
- Passive seismic: 3D velocity distribution
- Muography: Density distribution (2D or 3D)

The 2D (mapping, section) and 3D (voxel cube) structures, which define the distribution of data and physical parameters in the integrated database, can be transferred to the integrated database after appropriate coordinate transformations and format transformations. The physical rock properties with continuous distributions must be properly discretized (or categorized) taking into account the content of their geological information (Hall & McMullen, 2004).

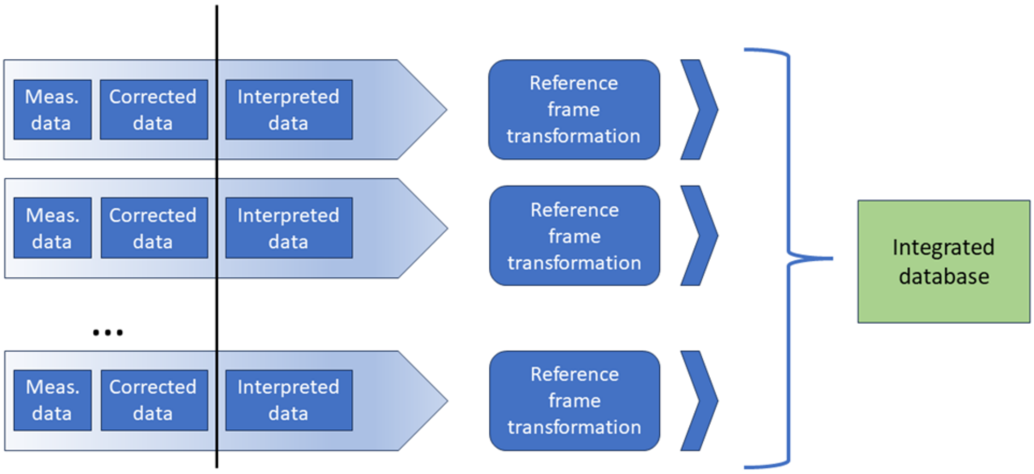


Figure 6. Overview of the data integration process

(Reference: Muon Solutions Oy)

The preliminary or temporary data and/or auxiliary information not required for geological interpretation (such as raw measurement data, sub-results, experimental parameters, etc.) are stored in the databases provided by the responsible institution (e.g., by those who conduct the actual measurements) while these are also a part of the AGEMERA project database.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

2.5 Integration of Surface and Subsurface Coordinate Systems: Insights from Pyhäsalmi

One way to build detailed topographic models is laser scanning. The data are usually built based on three-dimensional point-like data illustrating both the ground and objects on the ground as usually those are not easy to separate. Each data point is represented in some coordinate system, such as commonly used Cartesian coordinates (x, y, z) often used in the local coordinate systems.

One particularly interesting example is that of combining the open-source global topographic data, such as the open data of the National Land Survey of Finland (providing datasets and interfaces), and the underground local coordinate system, such as the block model of the Pyhäsalmi mine. The open laser scanning data of the National Land Survey of Finland (NLS) can be encapsulated as follows (link, accessed 20 Oct. 2023):

- Purpose: Laser scanning data are extracted for constructing topographic models
- Geographic location: entire Finland
- Reference system: ETRS89 / TM35FIN(E,N) (EPSG:3067) N2000 height (EPSG:3900)
- Spatial representation: Vector
- Data content: Laser scanning data

As an example, the topographical data are combined with those of the block model of the Pyhäsalmi mine. The idea is to match the underground block model to topographically characteristic targets that are clearly characteristic for both models. In this case the open pit is a rather clear choice. The results are shown in Figure 7 and suggest that the present approach may be an interesting way to connect the local underground and global coordinate system. Figure 7 consists of six parts. Upper left: The local topography (e.g., open pit where the solid rock is exposed, and loose soil has no effect on the surface model) in the NLS coordinate system. Upper right: The topography in the local coordinate system. Middle left: The local system (before the coordinate matching) and the NLS systems placed in the same coordinate system. Middle right: The rotated and shifted local coordinate system such that it corresponds to that the NLS system connecting them as one reference system. Lower left: The local system (after the coordinate matching) and the NLS system placed in the same coordinate system. Lower right: The difference between the two models in percentages together with elevation curves. Note that the refilling of the open pit may change the surface models and it is seldom taken into account in block models which concentrate on solid rocks rather than refilling. Therefore, one must be careful while choosing the reference points in the surface models.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

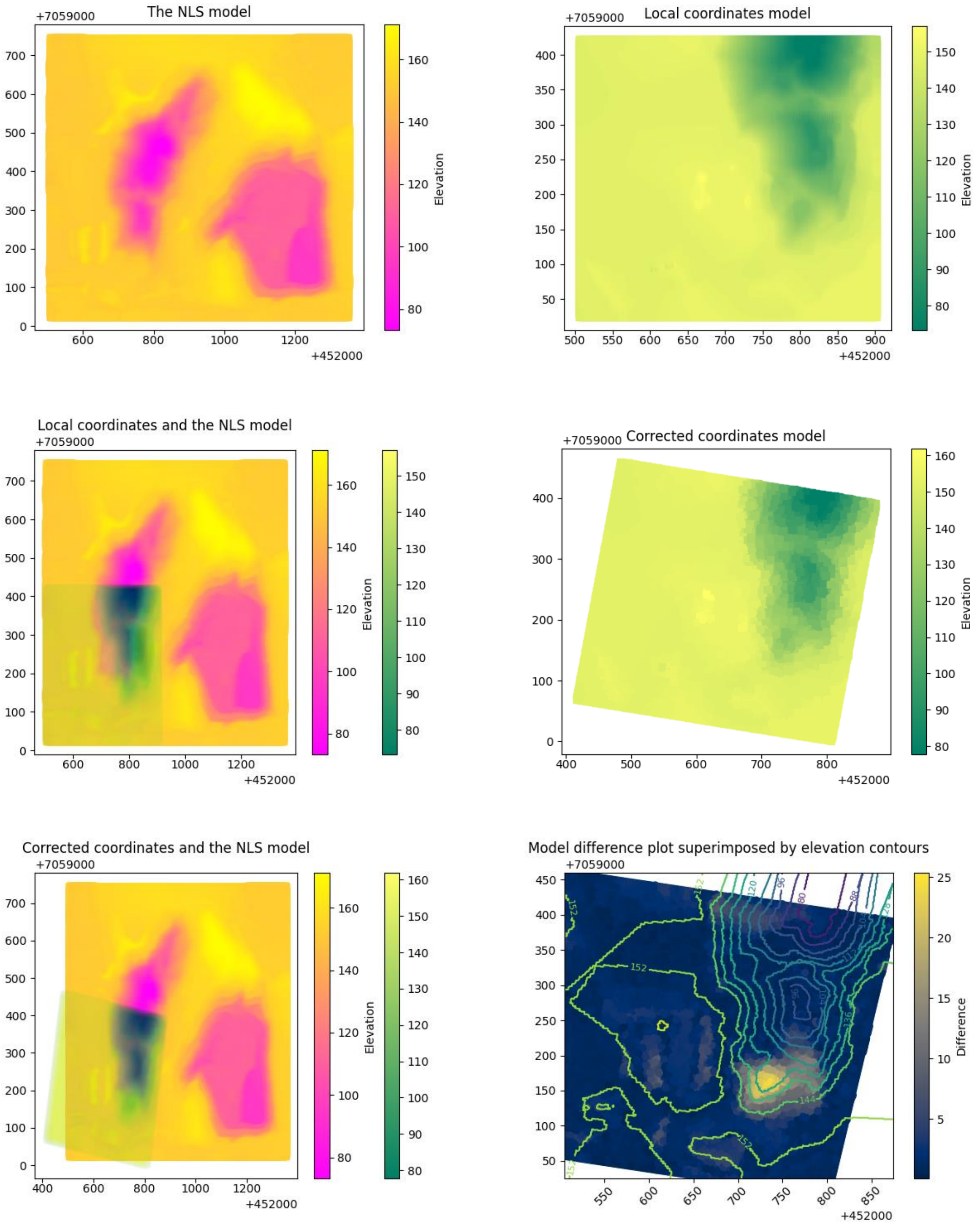


Figure 7. Matching the underground block model of the Pyhäsalmi mine and the surface topographical model provided by the National Land Survey of Finland (NLS) using the open pit as the reference characteristics. All axes are in the map coordinates provided by NLS (in metres using the ETRS-TM35FIN and geodetic datum of EUREF-FIN) in the global reference frame. Vertical axis is towards the North and horizontal towards the East. Elevations are in metres from sea level. For more details, see text

(Reference: Muon Solutions Oy)

The basic idea is simple: use the selected surfaces or surface points of two models, align the surfaces by rotating and shifting them both vertically and horizontally. Depending on the topographic characteristics and the selected area, the matching may be very accurate and, generally speaking, the more characteristics and the larger the selected area, better they match. However, it is worth noticing that, for example, the refilling of the open pit may change the surface models and it is seldom taken into account in the block models, which concentrate on solid rocks rather than refilling. Therefore, one must be careful while choosing the reference points in the surface models.

When the current underground coordinate system is not connected to any global coordinate system and there are at least some topographical characteristics, this method could provide a simple and straightforward tool to link these two systems together. Furthermore, if these data already are available, the method provides a cross-check whether the current underground coordinate system matches that of global and if not, what is the difference between those two (e.g., in percentage or metres). Therefore, this simple method may provide valuable information concerning the underground coordinate systems as particularly old mine maps may be rather inaccurate. The systematic errors in positions may be large because the possibilities for the accurate underground positioning are very limited. Therefore, we are investigating further both the method and its possible applications.

2.6 Data Storage

The main output data storage used by the AGEMERA platform is Amazon Simple Storage Service (S3), a cloud storage service provided by Amazon Web Services (AWS). It allows for storage and retrieval of different data types in a highly scalable and durable manner. S3 is accessible via web interface, command-line tools or various software development tools, to authorised users only, providing data security. It is designed for high availability and durability, making it a popular choice for data archiving and serving content for websites and applications, such as AGEMERA GUI.

On the other hand, the metadata, used in the product outputs, is automatically generated during the preparation of the task based on the specific parameters, where the terms, nomenclature and naming conventions of the individual data keys are standardised, and are stored in GeoJSON file format. All GeoJSON metadata can be viewed, searched and manipulated through PostGIS database.

Finally, the platform also employs OGC Web Map Service (WMS), GEO web service standards adopted by INSPIRE and ISO, for visualization of map stacks.

The main data storage for input data, received from partners/data providers is agreed upon in the project’s Teams folder (Data Repository), shared and accessible only to project members.

2.7 Data Accessibility

When it comes to data accessibility and security, AGEMERA platform offers a comprehensive solution. All output data is accessible and can be analysed conveniently through the platform's Graphical User Interface, known as AGEMERA GUI. The initial sign-in procedure grants access to the web-based GUI for everyone, but we've taken extra steps to ensure data security and control data visibility based on User Groups.

This means that data obtained from public and openly available sources remains visible to the general public, specifically to those who register on our platform. However, data providers or owners of proprietary data have the flexibility to limit the visibility of their data exclusively to consortium members, who are part of a pre-defined User Group. This decision regarding data visibility can be communicated at the same time as the data is provided to OPT/NET for ingestion into the platform.

Furthermore, it's essential to note that there is no two-way communication between the platform's frontend component (GUI) and backend component (OCLI). This design choice ensures that data cannot be downloaded from the platform's GUI; instead, all ingested data can only be visualised, adding an extra layer of security to safeguard sensitive information.

3. Data Categories and Dimensions

3.1 Introduction to Geospatial Data

Regarding geospatially referenced data, there are two primary categories of data: vector and raster data types. Vector data, the more prevalent form, typically constitutes the bulk of information processed in GIS software. It symbolises geographical entities as points, lines, or polygons. On the other hand, raster data encapsulates geographical information in a grid-like arrangement of cells, each holding a specific attribute value. Unlike the variable areas represented by polygons in vector data, each cell in a raster dataset covers a uniform spatial extent, known as its 'spatial resolution'.

Vector objects typically offer a more discrete and more accurate representation of complex geological, geochemical, and geophysical objects than raster objects. Advantages related to vector data include that the data may offer a high level of detail and accuracy, data are easily scalable without loss of quality, and additional information can be stored easily in attribute tables (e.g., metadata). Moreover, with vector-based data, one can represent complex spatial relationships and generate new data layers using various GIS software tools. Vector data may be computationally intensive due to their complexity, particularly when representing complex features. Also, vector data may not be ideal for representing continuous data like elevation, bathymetry or temperature. They are also sensitive to the scale of representation, which can introduce inconsistencies when integrating data from different sources or scales. High-detail vector data can result in large file sizes. Furthermore, manual digitisation processes (if needed) can introduce human errors, affecting the overall accuracy of the data set.

Rasterised objects often originate from a variety of sources, including satellite imagery, airborne surveys, and ground-based measurements. Old paper copy reports and publications, particularly those containing maps, charts, or other graphical representations, are frequently digitised to create rasterised data. This process involves scanning the paper documents and converting them into a grid of pixels, each with associated values representing different attributes such as colour, intensity, or elevation. These rasterised versions of historical or archival documents can then be integrated into modern Geographic Information Systems (GIS) for further analysis and interpretation. Rasterised data is well-suited for representing continuous spatial phenomena. Limitations related to raster data may include data incompleteness, resolution inconsistencies, and potential errors introduced during the rasterization process.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

3.2 Primary Data Categories

The process of data fusion in the AGEMERA project relies on a variety of primary data categories, each contributing unique and valuable insights into the geological, geochemical, and geophysical features of the selected CRM deposits. These primary data categories serve as the foundational elements upon which more complex analyses and interpretations are built. They are also the key inputs for the data integration process, which aims to amalgamate contrasting data types into a cohesive and comprehensive understanding of the geological features of each field trial area. This section delves into the specifics of these primary data categories.

3.2.1 Geological Data

Geological data form the cornerstone of any mineral exploration or geological research project. In the context of the AGEMERA project, these data are primarily collected in WP1 and serve multiple purposes, from identifying mineral compositions to understanding geological structures and formations. Geological data are often multidimensional as there are many different types of data one can collect. These include descriptive information on rock types and their mineral composition as well as numerical geochemical data on samples collected from outcrops. Structural geological data forms another important geological data type.

Rock Samples:

Rock samples are indispensable for obtaining direct, tangible information on the geological features of interest. These samples provide critical insights into mineral composition, texture, and other petrological attributes. Typically, they are collected from outcrops, drill cores, or through other subsurface sampling methods, such as trenching or channel sampling. Data obtained from studying rock samples can be numeric (e.g., whole-rock geochemical data, mineral compositional data) and with known coordinates, or it can be descriptive with or without exact coordinates (e.g., mineralogical data), depending on the scale and type of the analysis. In places, “data” is not only descriptive but also deeply based on personal experiences and expertise. Below is a summary of the most important data types provided by geologists.

Mineralogical Data: Mineralogical data are **predominantly two-dimensional** in nature, and newly acquired data are usually represented in a vector-based format. Older data can also be represented as raster images. The selection of specific methods is contingent upon the given objective and may evolve as the project progresses. Common methods include:

- **Microscopy:** Optical microscopy is used for the initial identification of mineral phases and textural relationships within the rock samples.
- **Scanning Electron Microscopy (SEM) and Cold Cathodoluminescence (Cold-CL) Imaging:** These techniques offer high-resolution imaging and are

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

particularly useful for studying mineral surfaces and the spatial arrangement of mineral phases.

- **Electron Probe Microanalysis (EPMA) and Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS):** These are employed for trace-element studies, providing quantitative elemental compositions of minerals.
- **X-Ray Diffraction (XRD) and X-Ray Fluorescence (XRF):** These methods can provide additional data on mineral phases and elemental compositions.
- **Fluid Inclusion Studies:** These offer insights into the fluids that have interacted with the rock, which can be critical for understanding mineralisation processes.
- **Geochronology:** This involves dating the rock samples to understand the timing of geological events.
- **Mineral Staining Techniques:** These techniques involve the application of various chemical reagents to rock samples in order to induce colour changes in specific minerals. The resultant staining patterns can provide valuable insights into mineral composition, grain size, and the spatial distribution of minerals within the rock matrix. Types of staining techniques include carbonate and sulphide staining. The various mineral staining techniques can be used, for example, to delineate ore bodies by a relatively rapid and cost-effective manner.

Geochemical Data: Geochemical data are **generally two-dimensional**, with new data typically being captured and represented in a vector-based format. Older data can also be in raster format. It is important to note that mineralogical studies, employing techniques such as microscopy, Scanning Electron Microscopy (SEM), and Electron Probe Microanalysis (EPMA), also contribute geochemical data. These data enhance our understanding of mineral composition and structure, complementing the geochemical analyses. Geochemical data are multi-dimensional in nature, encompassing a variety of chemical elements and compounds in different states and concentrations. These data can be quantitative and expressed in units like parts per million (ppm), weight percent (%), or molarity (M), or they can be qualitative, indicating the presence or absence of specific elements or compounds.

<u>Data Collection and Methodologies:</u>

Outcrop Mapping: Outcrop mapping holds a critical position within the AGEMERA project, particularly in the designated project areas of Bosnia-Herzegovina, Bulgaria, and Spain. The data gathered through this method serve multiple functions, including the collection of rock samples for geochemical and mineralogical studies as well as providing insights into structural or hydrothermal controls on ore deposits. Outcrop mapping yields **two-dimensional data** that delineate the spatial distribution of various rock types within a given outcrop. However, if an outcrop contains measurable structural geological fabrics, the data often gets a **third dimension** (see below).

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

Structural Geological Data: The methodologies for structural geological outcrop mapping can vary, encompassing both traditional (e.g., compass clinometer measurements) and innovative techniques (e.g., digital mapping technologies). Data obtained from structural measurements are typically three-dimensional in nature, capturing both the orientation and magnitude of geological features. These data sets include parameters such as dip and dip direction, plunge and trend, or strike and rake, which collectively offer a comprehensive spatial understanding of geological structures like faults, folds, and foliations. Structural geologists typically focus on several key parameters when conducting outcrop mapping. These include foliation, which provides insights into the planar fabric of rocks; fold axis, which helps in understanding the geometry and orientation of folded structures; and mineral lineation, which can indicate the direction of tectonic forces. Additionally, fault dipping directions and angles are meticulously recorded, as they offer crucial information on the movement history and stress regimes.

3.2.2 Geophysical Data

Geophysical data serve as a fundamental pillar in mineral exploration and geological research projects. These data fulfil a range of objectives, from delineating subsurface structures to identifying variations in physical properties that may indicate the presence of mineral deposits.

Conventional Geophysical Data:

Geophysical data are typically measured as a function of position either at individual points (point data), along more-or-less straight profile lines (line data) or along multiple adjacent profile lines covering a larger survey area (area data). In addition to position, the data are often a function of some other property of the measurement system, such as frequency, time channel, transmitter-receiver offset, etc. The line azimuth (heading angle from north) is defined based on known or assumed geological strike direction of the survey area. This makes it more justified to interpolate the data over entire survey area to fill the gaps between lines, since the line separation is usually bigger than the data sampling distance along the profile lines. The line separation and point sampling are selected based on the geological complexity of the area and the physical properties of the measured geophysical field quantity. For example, gravity surveys can be made using sparse sampling (100-500 m) if we already know that the depth to the anomalous targets is big, but denser sampling (10-50 m) is needed if targets are located at shallow depth.

Some geophysical data are merely processed and prepared in such a way that they can be utilized for qualitative analysis. For example, radiometric data can be processed to yield apparent concentrations of potassium, uranium and thorium in addition to the total gamma-ray intensity. Depending on data coverage the results are presented as maps. After data processing, seismic data and ground penetrating radar (GPR) data are presented as vertical depth-sections (where the depth is not absolute but based on assumptions).

Whenever it is possible, geophysical data are interpreted using numerical inversion of 1D, 2D or 2D models. Depending on the geological problem, type of data and available software, the processed inversion results can be presented as 2D horizontal maps, which can represent (petro-) properties of the subsurface on various layers or depth sections. Profile data are often presented as vertical 2D cross-sections, and multiple profiles can be interpolated to 3D mesh (voxel) models that, in turn, can be visualized in various ways (e.g., volumetric views with transpareny, isosurfaces, ortho- and obliqueimages). Some data can even be visualized as vectors, where the length (or color) and direction represent field intensity and direction.

Seismic Techniques for Structural Data:

In addition to traditional geophysical methods, seismic techniques based on environmental seismic noise are employed in AGEMERA to offer structural information of the shallow subsurface up to depths of 2-3 km. These techniques also provide 2D and 3D distribution data on physical properties that can reveal significant structures controlling the mineral system.

3.2.3 Satellite Data

Satellite data sources are of great importance in the mining industry for many reasons. First, they provide a comprehensive and real-time view of the Earth's surface, enabling mining companies and interested parties to identify potential mineral deposits with remarkable precision during the prospecting phase. This reduces the time and resources spent on ground surveys and minimizes the environmental impact associated with exploratory drilling. Second, satellite data continues to prove invaluable in environmental monitoring throughout the lifecycle of mining operations. From the initial excavation stages to the eventual site closure, satellite data allow for the continuous environmental and change monitoring. This not only ensures compliance with environmental regulations but also facilitates the timely implementation of mitigation measures to prevent or minimize harm. Third, by monitoring factors such as weather conditions, infrastructure development, and transportation networks, mining companies can make informed decisions about production schedules and logistics.

As highlighted by the European Union Agency for the Space Programme (EUSPA), the role of EU Space in the mining sector is predicted to become a central driver in achieving the objectives laid out in the European Critical Raw Materials Act, particularly in securing a reliable supply of critical raw materials. It enhances exploration efficiency, supports responsible environmental impact, and contributes significantly to the broader goal of achieving resource independence and economic resilience in the critical raw materials sector.

3.3 Data Dimensions

3.3.1 2D Data Types

Two-dimensional (2D) data types are integral to the process of geological and geophysical data fusion. These data types offer spatial information in a planar format, capturing variations along two axes — typically latitude and longitude, X and Y, or Northing and Easting coordinates. The data can be broadly categorised into vector and raster types, each with its own set of characteristics and applications. This section elaborates on the various 2D data types utilised in the AGEMERA project, categorised by their spatial characteristics and data format.

2D Vector data:

- **Point Objects:** These include, for example, geodetic control points, measuring points, borehole collars, sample locations, outcrop locations, specific mineral observation points, geochemical analysis points, and photograph points. *These features are essential for ground-truthing and spatially locating various geological and geophysical phenomena.*

- **Linear Objects:** This category encompasses, for example, measurement lines, geological fault lines and drilling trajectories projected onto the ground surface. If not enclosed features, linear data may also comprise data layers such as rock type boundaries, hydrothermal alteration envelopes, ore body boundaries and exploration excavations. *These features are vital for delineating geological structures and guiding exploration and mining activities.*

- **Polygonal Objects:** These include, for example, given company areas, mine layouts, site infrastructure, research areas, and environmentally sensitive zones. Such data layers as boundaries of rock types, outcrop outlines, hydrothermal alteration envelopes, ore bodies, and exploration excavations are polygonal data if they form entirely enclosed features. In addition, 2D geophysical measurement data is typically polygonal. *Polygonal objects are used for advanced spatial analyses and for defining areas with multiple attributes or constraints.*

- **Topography:** This category includes elevation profiles, slope gradients, and terrain roughness. *Such data are indispensable for understanding surface morphology and are often sourced from digital elevation models (DEMs) and topographic maps.* It should be noted that topographic data are most frequently represented in raster format.

2D Raster data:

- **Topography:** The raster-form topographic data category includes elevation profiles, slope gradients, and terrain roughness. Such data are indispensable for understanding surface morphology and are often sourced from digital

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

elevation models (DEMs) and topographic maps. It should be noted that while topographic data are commonly represented in raster format, they can also be represented in vector format.

- **Geological, Geochemical and Geophysical Objects:** These types of rasterised objects may be newly generated data sets, but more commonly, they are digitised versions of historical publications, paper-only reports, etc.
    - 2D geological raster objects can include, for example, lithological and alteration maps with or without information on point and line type data (e.g., outcrop locations, specific mineral observation points, fault lines). *Lithological maps, for instance, provide a spatial representation of rock types at the Earth's surface or subsurface, which is crucial for understanding the geological history of a region. Fault lines captured in raster data can be invaluable for improving understanding of genetic processes that led to the formation of a mineral deposit.*
    - 2D geochemical objects in raster data often comprise elemental distribution maps, mineral abundance maps, and isotopic ratio maps. Elemental distribution maps can reveal the concentration of specific elements like copper or cobalt across a geographical area, which is *vital for mineral exploration and genetic modelling*. Mineral abundance maps *can indicate the presence of economically valuable minerals or hydrothermal alteration*. Isotopic ratio maps *can provide insights into geological processes* such as hydrothermal fluid movements and mantle-crust interactions.
    - 2D geophysical objects in raster data can encompass magnetic anomaly maps, electromagnetic anomaly maps, gravity anomaly maps, and seismic reflection profiles. Different types of geophysical anomaly maps are instrumental in identifying subsurface structures, which is essential for mineral exploration and understanding local geology (e.g., sedimentary basin architecture).

3.3.2 3D Data Types

The 3D data (d(x,y,z)) in the merged database are derived partly from geophysical measurements and partly from geological model elements which are used as an input to build the 3D models. On the geophysics side, the tomographic methods such as passive seismics and muographic measurements produce such data (velocity distribution and rock density distribution, respectively). However, it is worth noticing that there are also other methods that may contribute. For example, in terms of electrical methods the 3D data can also be produced using the electrical resistivity tomography (ERT) and electromagnetic (EM) methods, and for the latter methods, it is possible to define a frequency dependent pseudo depth for the depth estimates.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

In the AGEMERA project, the treatment of the 3D (x, y, z) frame data consists of a transformation of the surface projection of the data system, i.e., (x, y) coordinates in the local surface projection and depth (z) in the local reference frame, and format transformations leading to the integrated 3D database. This is illustrated in Figure 8 which shows the data handling procedure from the local frame to the integrated 3D database.

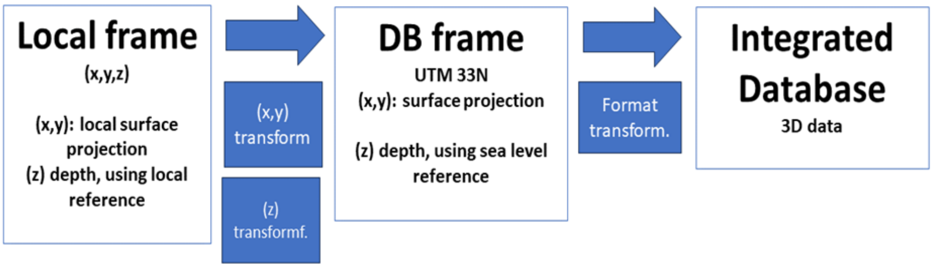

Local frame
(x,y,z)
(x,y): local surface projection
(z) depth, using local reference
(x,y) transform
(z) transformf.
DB frame
UTM 33N
(x,y): surface projection
(z) depth, using sea level reference
Format transform.
Integrated Database
3D data

Figure 8. 3D data handling from the local frame to the integrated 3D database via further format transformations in the DB frame

(Reference: Muon Solutions Oy)

In the AGEMERA project, searchability is provided by the link to the integrated database map projection. The 3D data are discretized both in space and value. In space, the physical data obtained from the measurement is represented by a voxel structure adapted to the resolution of the given method. The voxels are defined in a Cartesian system, and in the first step the surface projection of voxel volume is related to the local map projection. This is illustrated in Figure 8 which shows a schematic view of the surface projection and the underground voxel system.

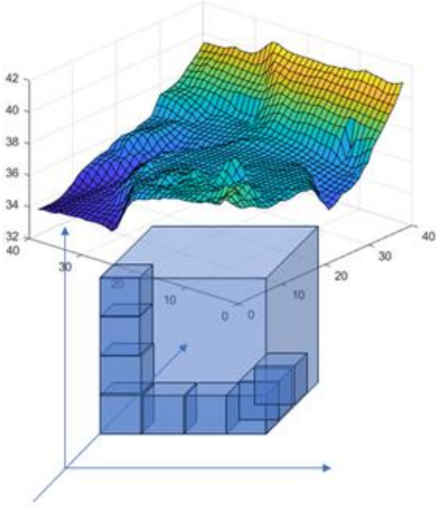


Figure 9. Schematic view of the surface projection and the underground voxel system

(Reference: Muon Solutions Oy)

In the database each voxel is characterized by a particular rock physical property or a particular geological property. Furthermore, although rather trivial, it is worth noticing that the transformation of the local surface projection also transforms the (x,y) coordinates of the voxels.

In the integrated database, the layers associated with the voxels are represented by 15 discrete values and this number is fixed. Accordingly, the values need to be categorized before being transferred to the integrated database. The categories should be constructed taking into account the geological information. It should also be borne in mind, knowing the variance of the method and the estimate, which rock categories can reliably be separated. As a final step only the transformed voxel coordinates and discretized physical values must be transferred to the integrated database. The (voxelized) 3D data also include metadata files with the recorded parameters used to generate the data.

3.4 Metadata

Metadata is descriptive information related to the data structures stored in the Integrated Data Base (IDB) that cannot directly be retrieved from the database. It is categorised, but not necessarily structured data. Metadata can be accessed through the identifier of the stored data structure. These identifiers can be, or related to:

- Parameters for its creation
- History of versions/processing status
- Reliability information
- Status of checks (validation)

Metadata can also be used to identify the source data of the data structure in the integrated database. The source data are stored in the database by the institution responsible for the creation of the data. These metadata allow IDB users to view the detailed history of the structure stored in the IDB. To avoid mistakes. the metadata files cannot be modified by the users of the integrated database. Metadata are divided in different categories which should be stored in a separate file identified by the same basic identifier. These identifiers are, for example:

- Measurement Parameters (3.4.1)
- Parameters of Data Processing (may apply to multiple measurements processed together) (3.4.2)
- Parameters of Interpretation (3.4.3)
- Coordinate Systems (3.4.4)
- Status and Reliability Information (for the data structure stored in the IDB) (3.4.5)
- Parameters of Transformations Required for IDB Input (3.4.6)
- Error Metrics (3.4.7)
- Validation and Verification Parameters (3.4.8)
- Audit trail (3.4.9)

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

It is essential to emphasize that the most important aspect is to record and note down all parameters necessary for the precise repetition of the measurement as the ability to replicate the measurement and reproduce its results are crucial in all scientific research, manufacturing processes, or any measurement activity in general. Thus, a precise and comprehensive recording of measurement parameters ensures that assessments are reliable and evaluable.

3.4.1 Measurement Parameters

The measurement parameters file contains all parameters relevant to the IDB user while referring to the measurements. These parameters can also be used to retrieve (with the involvement of the responsible institution) the raw measurements and all related additional parameters.

It contains the identifiers, type, date and time of the measurements which were used to create the data structure stored in the IDB. It also includes the coordinates needed to localise the measurements in detail, identifiers for involved institutions and persons responsible for the measurements, including identifiers of all employed instruments. It also lists the file formats of raw measurements, and all findings on the quality of the measurement. In other words, all detailed measurement reports, set-up parameters for all employed instruments, graphs, etc. are available in the database of the responsible institution and can be searched using the measurement identifiers.

3.4.2 Parameters of Data Processing

This file contains relevant parameters to the IDB user in all data processing. These parameters can be used to retrieve the partial results of the processing (with the involvement of the processing institution in case the access is limited).

The file also includes the identifiers of the measurements involved in the processing, the date of processing, the processing version number, the main processing steps (method and basic parameters defining each step), the identifiers of the institution and persons performing the processing, the identifiers of the processing software, and the data structure the data are stored in the IDB. It also includes comments concerning the quality of the processing. Detailed processing data can be accessed by the processing version number in the database of the processing institution.

The discretisation of the measured domain (voxels, layers) plays an important role in several processing steps (e.g., inversion). Information on the voxel geometry must also be stored.

It is worth noticing that accurate and precise documentation of all data processing steps, methods, software used in the data processing as well as data sharing procedures contribute to the reliability of research and the verifiability of results.

3.4.3 Parameters of Interpretation

This file contains all parameters relevant to the IDB user's interpretation of the processed data system. These parameters can also be used to retrieve (with the involvement of the interpreting institution) the partial results of the interpretation and the documents relating to the analysis procedures.

The file also includes the identifiers of the measurements involved in the interpretation process, the identifiers of the processed data structure, the interpretation version number, the purpose of the interpretation, the main assumptions, the parameters of the model used, the parameters for discretising the interpretation result, and the data structure the data are stored in the IDB.

It is also important to identify the institution and persons performing the interpretation as well as the interpretation software in detail and all possible comments concerning the quality of the interpretation (cf. Error Metrics below).

3.4.4 Coordinate Systems

Most geological information requires at least some kind of coordinate system so that the data can be localised in terms of position. While different coordinate systems are already explained in detail (see, for example, Sect. 2.2) those are such important parameters also from the metadata point of view that they are worth repeating.

The data formats and coordinate systems supported by the AGEMERA project are:

- 2D data:
    - File format: GeoTIFF (for raster data), KML and GeoJSON (for vector data)
    - Coordinate system: EPSG 4326 (another option: the UTM coordinate system)
- 3D data:
    - File format: NetCDF
    - Coordinate system:
        - EPSG 4979 (ellipsoidal 3D coordinate reference system for 3D objects on the surface of the Earth),
        - EPSG 4978 (geocentric 3D coordinate reference system for the objects above, below and in orbit around the Earth), or
        - EPSG 5773, EPSG 5798 and EPSG 3855, which are vertical datum for EGM96, EGM 84 and EGM2008, respectively. Note that these represent only the vertical axis with zero-height surface resulting from the application of the EGM96/EGM84/EGM2008 geoid model to the WGS 84 ellipsoid.

3.4.5 Status and Reliability Information

This file contains metadata describing the current state of a given data structure in the IDB. The status data (Inadequate, Temporal, Test Data, Final, Verified,

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

Validated) can also be checked by the user for each element of the data structure creation (for each version and/or the latest version).

3.4.6 Parameters of Transformations Required for IDB Input

A critical element for data storage in the IDB is the appropriate coordinate and format transformation of the incoming data (see, for example, Chapter 2). The parameters of this transformation should be stored as meta-parameters in the parameter transformation file.

The file contains the local coordinate system used by the responsible institution that is used for generating the data (for measurement and evaluation). It also includes the coordinate transformation used for the IDB input, its parameters (e.g., used software), and the identifiers of the persons performing the coordinate transformations. The quality control parameters of the transformation must also be stored.

Yet another set of parameters to be stored are parameters for possible rescaling of data with different resolutions (e.g., pixel or voxel sizes) as those require different adjustments for the discrete data sets.

3.4.7 Error Metrics

All measurements contain sources of errors and therefore error analysis is an essential task while examining the reliability of the results. It is also a well-known fact that the error analyses/reliability checks are one of the most challenging parts of all measurements.

In both data fusion and geological interpretation, it is important to specify the uncertainties and information contents of the used data. These are usually expressed by means of statistical models associated with data collection (e.g., measurement), data processing and estimation (e.g., data analysis), often with parameters of assumed error probability distributions.

In general practice, a measure is associated with the inputs to geological modelling, which through the process of error propagation, usually include the error metrics of the elements incorporated in the previous measurement and processing phase. Thus, the error metric is the standard deviation of the data for the given (or assumed) distribution. The standard deviation with the assumed probability distribution can also be used to define confidence levels and intervals.

In case of noise and other interfering processes, a variance (or the squared standard deviation or the average of the squared differences from the mean) can also be used to examine the question of sensitivity (or detectability). For each measurement and the processing of measurement, the estimated standard deviation of the result or sub-result is calculated allowing to track the propagations of errors individually.

This is important as in the data evaluation process the measured parameters are often correlated, e.g., due to overestimation or incomplete design of

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

measurement. Therefore, it is worthwhile to examine the parameter covariance matrix (i.e., a direct measure for the correlation) as suspiciously high correlations indicates inadequate model construction. Thus, the covariance is also important for the correct handling of error propagation. However, due to its generally large size (if dealing with the results of a tomographic problem), it is not expected that the covariance matrix can be included in the integrated database. Therefore, it may be better to select the data host institution to be responsible for taking care of this storage which may also be a part of the quality control.

In some cases, the selected methods may contain elements that introduce bias in the estimation results (particularly if the methods use regularization). In such cases, it is useful to provide upper bound estimates for these elements. Furthermore, random errors (usually uniformly distributed errors) are generated by discretisations and roundings and those should be taken into account when specifying the error rates.

3.4.8 Validation and Verification Parameters

Validation and verification are also key processes, and it is advisable to record its main parameters separately. This approach provides an overview of the control (where necessary validation) points in the whole process of setting up the IDB structure. It is also important that the user can always track the status of the validation of any given structure.

In addition to status information, the check parameters (e.g., time, person(s) performing the check/validation, comments concerning the check, comments on the corrections) must be available.

3.4.9 Audit Trail

This file provides an overview of the history of the IDB structure (timestamped data processing and control steps), allowing the reader to follow the whole process.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

4. Data Types

4.1 Lithological, Structural Geological and Mineralogical Data

Within the framework of Task 1.2, field research focuses on outcrop mapping in selected CRM deposits. Structural geological mapping activities are designed to produce input to the computer-based modelling endeavours under Task 1.4. The overarching objective of these field studies is not merely confined to the examination of rocks in the vicinity of the selected mines and mineral occurrences. It also involves a detailed analysis of key lithologies and outcrops, which may contain invaluable geological information essential to the grades and locations of CRMs.

Lithological Data: Lithological data is collected by applying the various techniques associated to outcrop mapping. These include typically **two-dimensional data** concerning rock types (e.g., rock names) in a given outcrop.

Structural Geological Data: Data collected during field campaigns or digitized from pre-existing geological maps are typically three-dimensional in character and consist of fault planes, kinematic indicators, bedding dips. Structural data are typically georeferenced points with attributes (e.g., dip and dip azimuth or trend and plunge). These data are then interpolated and used to build 2D and 3D structural models for selected test sites, namely Bosnia-Herzegovina/Croatia and Mina Concepcion, Spain. The models will be built using MOVE software (developed and commercialized by Petroleum Experts). Exports from the models will be lines or geological surfaces in xyz format.

Mineralogical Data: Mineralogical studies on rock samples are conducted using a range of advanced laboratory methods and analytical tools. Among the methods so-far used, or currently considered, are:

- ***Microscopy:*** Optical microscopy has been used for the initial identification of mineral phases and textural relationships within the rock samples. Most descriptive pictures are saved in jpeg file.
- ***Electron Probe Microanalysis (EPMA) and Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS):*** in AGEMERA, EPMA was employed for trace-element studies and detection of the accessory minerals, providing quantitative elemental compositions of minerals. Qualitative chemical composition of accessory minerals is presented in the jpeg file. Quantitative data one the chemical composition of the selected minerals is presented in the tabular form (.xlsx). LA-ICP-MS was used for bulk rock analysis of trace elements. Data are presented in tabular form (.xlsx).
- ***X-Ray Diffraction (XRD) and X-Ray Fluorescence (XRF):*** These methods can provide additional data on mineral phases and elemental compositions. XRD was used simultaneously with microscopy as additional method for identifying very small mineral phases. Data are

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

presented as diffractogram in the jpeg file. XRF is used for obtaining the major chemical elements composition of the bulk rock samples. Data are presented in the tabular form (.xlsx).

- ***Geochronology:*** This involves dating the rock samples to understand the timing of geological events. In AGEMERA, this method has so far been applied on dating bauxites from Bosnia-Herzegovina. This method implies LA-ICP-MS on the zircon separates. Results are presented as diagram on the jpeg file.

The primary aim of these mineralogical studies is to furnish missing geochemical and mineral-chemical data. This information is crucial for the computer-based modelling activities under Task 1.4 (T1.4) and for enhancing the mineral system models for CRM deposits under Task 1.5 (T1.5). Additionally, these mineralogical studies serve to establish a more robust link between geochemistry and mineralogy with geophysics, as well as with the results emanating from the innovative new methods being applied in WP3 at the target sites. They are, of course, useful also in data fusion.

Geochemical Data: Geochemical analyses serve as a foundational element in the AGEMERA project, offering a detailed chemical data from the samples collected from the field. These analyses aim to quantify various elements and compounds, thereby shedding light on mineralisation processes and the presence of CRMs. While it is unlikely that all methods will be utilised in the AGEMERA project, which is still ongoing at the time of writing, the types of geochemical data generated can include:

- ***Elemental Composition:*** Quantification of individual elements within samples is commonly achieved through techniques like Inductively Coupled Plasma Mass Spectrometry (ICP-MS) or Atomic Absorption Spectroscopy (AAS).
- ***Compound Analysis:*** This category focuses on the identification and quantification of specific chemical compounds, such as oxides, sulphides, or carbonates.
- ***Isotopic Ratios:*** These studies offer insights into the origin and evolutionary history of mineral deposits by analysing isotopic ratios of specific elements.
- ***Trace Elements:*** Elements present in low concentrations are of particular interest, especially in the context of CRMs. Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS) is often employed for these studies.
- ***pH and Redox Conditions:*** Data on acidity or alkalinity and oxidation-reduction conditions are crucial for understanding mineral solubility and mobility.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

- *Organic Geochemistry:* This involves the study of organic compounds like hydrocarbons or organic acids, which may influence mineralisation processes.
- *Fluid Inclusions:* Analysing the chemical composition of fluid inclusions can provide critical insights into the fluids that have interacted with geological formations, aiding in the understanding of mineralisation processes.

It should be noted that certain types of chemical data may present challenges for integration in data fusion processes. For instance, fluid inclusion data can be particularly complex to incorporate due to its intricate nature.

4.2 Data Types in Drone Geophysics

4.2.1 Magnetic Surveys

The drone magnetic data are processed with Radai's own software using equivalent layer modelling method. The results are presented as GeoTIFFs that are clipped inside the survey area. In addition to total intensity of the magnetic field (TMI), derivatives, such as 1st and 2nd vertical derivative, horizontal gradient, tilt gradient and total gradient are used. The results are also given for pole reduced (RTP) case.

In some cases, 3D inversion is made using the vector magnetic data (individual XYZ components) and the results (3D mesh model of magnetic susceptibility or magnetization) are saved in a column formatted text files to be visualized using 3rd party software.

4.2.2 Electromagnetic Surveys

The EM data are processed with Radai's own software using a 1D model and joint inversion of EM and static magnetic field. The results are interpolated and stored as grids in GeoTIFF format. Once available, 3D inversion of EM data can be performed, and the results (3D mesh model of electric resistivity) are saved in a column formatted text file to be visualized using 3rd party software.

4.2.3 Radiometric Surveys

The drone radiometric data of Radai are processed to yield (gamma ray) counts per second and then interpolated and saved as GeoTIFF grid files.

4.3 Data Types in Passive Seismics

In passive seismics, we apply two different approaches to study the subsurface: seismic noise interferometry and estimations of Rayleigh wave ellipticity. Both approaches will be the same at the two selected test sites for this project.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

However, the analysis of the obtained results will be tailored to the specific requirements of each site, resulting in different outcomes.

4.3.1 Mina Concepción (Sandfire MATSA)

The main objective of applying passive seismic techniques at this site is imaging the shallow subsurface in the vicinity of Mina Concepción from both a lithological and structural perspective. With these techniques, we can extract different seismic observables that will be compared with the geological structure of the subsurface. In this way, we will be able to characterize the medium performing 2D and 3D seismic models of the study area.

Reflectivity zero-offset profiles (autocorrelations):
Seismic noise interferometry allows us to retrieve zero-offset reflectivity profiles while working with the single-station approach (Schimmel et al, 2011). In other words, this methodology allows us to extract the reflectivity pattern occurring directly beneath each seismic station. This approach enables the identification of geological discontinuities such as lithological contacts or faults in the medium under study.

The first type of resulting data is seismic records in conventional seismic formats such as mini SEED and SEGY, similar to the seismic records obtained in controlled-source seismic experiments. This means we will deliver a seismic trace for each seismic station deployed in the area. Integrating the results of all the seismic stations is possible to build 3D maps and identify robust reflections that will be associated with geological discontinuities in the medium (Romero and Schimmel, 2018). Then, a second type of data can be surfaces of contact, similarly obtained to the 2D and 3D controlled-source seismic profiles. This type of data will mainly consist of XY grids showing the position of characterized surfaces.

3D S-wave velocity models (cross-correlations):
The second application of seismic interferometry is to perform ambient noise tomography (ANT), which pursue the extraction of seismic signal from seismic noise to provide a final 3D S-wave velocity model of the subsurface. This technique allows the identification and the extraction of the waves traveling between station pairs as body and/or surface waves. Here, we will focus on the extraction of surface waves, and particularly Rayleigh waves. In a dispersive medium, the velocity of Rayleigh waves varies with frequency and therefore they will have different penetration depths. Estimating the dispersion curves for all the station pairs in the area, it is possible to build 2D Rayleigh wave velocity maps for different frequencies. Defining a grid in the study area, we can use all the 2D velocity maps to extract local dispersion curves as function of grid point. The inversion of each local dispersion curves results into a seismic velocity-depth profile for that grid point. The inversion of all local dispersion curves for each grid point thus permits to build a 3D velocity model. The final model is thus a 3D grid of XYZ coordinates and an attributed S-wave velocity value for each position defined by the XYZ coordinates (Nuñez et al., 2020).

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

Analysis of Rayleigh ellipticity:

This application is based on extracting Rayleigh waves directly from the seismic noise records. Rayleigh waves are characterized for having a particle motion that is polarized along a vertical ellipse, typically with retrograde motion at shallow depths. Once Rayleigh waves are extracted, it is possible to compute the ratio between the vertical and the horizontal axes of the particle motion as a function of the frequency. Since ellipticity only depends on the local structure beneath the station, and the inversion of these ellipticity curves depends on the frequencies, the method allows the building of 1D S-wave velocity profiles just beneath each seismic station (Berbellini, et al., 2019). The outcome format will be text files describing the S-wave profiles, including the velocities and depth interval. Same as before, the dense seismic network deployed in Mina Concepción will allow us to provide a smooth 3D velocity model (Jones et al., 2021) that will be delivered as an XYZ file with the S-wave velocity measurement at each node of the interpolated mesh.

4.3.2 Lubin Mine

In this case, we plan to monitor the subsurface's structural and mechanical properties with seismic noise interferometry. We will perform this task with auto- and cross-correlations of the seismic noise records in the area. The time evolution of the correlations can be characterized by two observables: the waveform similarity and the seismic velocity. Same as before, we can integrate the results corresponding to each station pair and build 2D maps of those observables varying with time.

The outcome format that will be delivered consists of in XYZ file with the estimations of waveform similarity and velocity changes for each day within the studied time span. In the case of different frequency range analysis, a fifth column is provided, indicating the depth location of the variation.

The application of the Rayleigh-wave ellipticity for time-lapse monitoring will be summarised by a similarity index. This index will indirectly quantify the variations in the subsoil velocities by applying a cross-correlation between the hourly-averaged ellipticity curves and a chosen reference. The curve that will serve as a reference will be built based on the average of the most stable period indicated by the dV/V variations. The final format will consist of a table of six columns containing the three spatial coordinates, the similarity index, the center frequency of the band analysed and the time of reference.

4.4 Data Types in Ground Geophysics

Conventional ground geophysical surveys are conducted in Bosnia and Herzegovina at various microsites to provide additional subsurface data to constrain geological models. We use different methods: electrical tomography to

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

determine the resistivity distribution in the subsurface, seismic refraction method to determine the differences in seismic velocities using the refraction of body waves in the subsurface, the electromagnetic method to determine the electrical properties of the subsurface by inducing EM energy into the subsurface and measuring the response of earth materials, and magnetic method to determine the differences in magnetic susceptibility (for magnetic minerals).

4.4.1 Electrical Resistivity Tomography (ERT)

Electrical resistivity data are processed using licenced and/or open-source software. Profile data are presented as GeoTIFFs or inverted models are stored in column-formatted text files that can be visualised with third-party software. This type of data consists mainly of 2D grids indicating electrical resistivities. In some cases, 3D inversion can be performed using multiple profiles, and the results (3D mesh model of resistivity) provide a smooth 3D resistivity model delivered as an XYZ file with the resistivity value at each node of the interpolated mesh.

4.4.2 Seismic refraction survey

Refraction seismic data are stored in SEG2 seismic format and processed using licenced and/or open-source software. Processed and interpreted data are presented as GeoTIFFs or 2D velocity models can be stored in column-formatted text files that can be visualised with third-party software. This type of data consists mainly of 2D grids indicating seismic velocities.

4.4.3 Controlled-source Audio-frequency Magnetotelluric (CSAMT)

The method allows the generation of 1D resistivity profiles directly under each magnetotelluric station. The modelled data can be presented in different forms depending on the arrangement of the stations. The MT soundings are collected along the profiles of interest and processed as 2D models so that the results can be presented as cross sections in GeoTIFF format.

4.4.4 Magnetic survey

The total intensity of the magnetic field is measured along profiles to investigate the possible presence of magnetic minerals in the subsurface, and the data are presented in the form of 2D diagrams (profiles).

4.5 Data Types in Cosmic-Ray Muography

Muography is employing cosmic-ray induced **muon particles** that are produced in the upper part of atmosphere (mainly at an altitude of 15 – 30 km). The produced muons can have very high energies enabling them to penetrate deep underground (down to kilometres of solid rock). As muography employs natural background radiation, there is no need for artificial radiation sources, but the data are recorded using muon detectors that detect the signals of passing muon particles (Holma, 2023a). The energies of most cosmic-ray induced muons is so high that they easily penetrate deep underground, and the depth is dependent

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

both on their initial energy and material thickness they pass through (the higher the initial energy and lower the material density, deeper the muons can penetrate) (Holma, 2023b).

In the mining applications, the registration of cosmic-origin muons is primarily aimed at an approximate reconstruction of the density distribution of the screened rock mass above the detector system. If the muons are detected from appropriate detector positions and with sufficient directional resolution, tomographic (3D) rock density reconstruction can be performed (Lesparre et al., 2010; Balazs et al., 2023).

Muography in 2D and 3D is based on muon **tracking,** where the path of each muon through the detector system is traced in detail. This provides direction-dependent data where the zenith (theta) and azimuth (phi) angles for each muon are binned in suitable angular bins according to the detector's angular resolution. So, the basic data for each muon can be described as a vector **r** having one known position (e.g., in Cartesian coordinates (x, y, z)) and two angles (e.g., zenith θ and azimuth ϕ). Thus, the data for each muon consist of five parameters: three positions and two angles, **r** = **r** (x, y, z; θ, ϕ). These data are usually recorded using a position sensitive muon detector where the position of the detector and the position of passing muon provides the (x, y, z) coordinates. The angular data are detected using some tracking algorithms providing two angles (θ, ϕ) for the muon initial direction. These two sets of data are enough for the unambiguous muon path through the matter. However, it is often useful to also record the time (t) of the passing muon track (i.e., an event time or time stamp) as time-dependent data are or can be used in many applications. Thus, the very basic muon data consists of six parameters: **r** = **r** (x, y, z; θ, ϕ; t) for each muon.

If the number of available muon detector positions is one, the result is 2D radiography. This kind of image shows the projection of the average density distribution in 2D. Therefore, it lacks the information concerning the distance to the anomalies while it is a proper description on the 2D density distribution. Perhaps the biggest weakness in this kind of image is that a small, dense object nearby looks the same as the large, less dense object from far away. In other words, the image is not unambiguous. However, it is worth noticing that all 3D imaging procedures start with 2D data handling.

If the number of available muon detector positions is more than one (i.e., stereo with ability to 3D), the result is 3D tomography. This kind of image tells the real 3D density distribution of the volume of interest within some accuracy. The latter depends on the number of detector positions, but also on the complexity of the anomaly. If the anomaly is complex, one needs several, if not many, different detector positions to optimise the 3D image of the anomaly of interest. It is worth noticing that the optimisation is not always easy but can be simplified either by careful studies or simulations before conducting the actual measurements or by using extra detector points to cover positions 'shadowed' by the complex parts of the anomaly (Varga et al., 2020).

The basic parameter in muography is the **muon counting rate** (or **muon flux**). This rate can be extracted either as directional or total. While the total is simply total (for the total muon flux, see below), the direction can have different angular resolutions, or different areas of the sky the muons are coming from. If the muon detection system allows **angular distribution**, or the system has any angular resolution, the muon data can provide direct directional information concerning the density of the volume of interest. In this method the muon counting rates (or muon flux or the number of muons as a function of time) are sorted and recorded in discrete angle bins. The discrete data are inputs for calculation of the directional **muon flux distribution** consisting of muons with all energies (energy-integrated) but only from the very limited direction. The total muon counting rate can also be used to estimate its statistical error (**variance**) and this accuracy information are recorded together with the measured data (Lesparre et al., 2010).

The muon counting rate (or muon flux) can be compared with the reference muon flux distribution at the surface, and the integrated rock density (or **density length**) for a possible muon trajectory can be calculated based on the attenuation in the muon flux (either directional or total). The average rock density (or **apparent density**) for each angle bin can be calculated using the density length and the distance the muon travels through the rock (the muon path length). While the density length is calculated using the muon attenuation (the deeper the less muons are detected) the muon path length can be extracted using the digital model of the surface and the position the muon detector is placed underground. As a result, the ratio between these two is the average rock density (apparent density).

The data from the muographic measurements and associated metadata are recorded and moved to the database where each measurement (RUN) has an individual sequence number which identifies the angular data associated with the measurement (including both measured raw data and transformed measured data) and the linked metadata containing parameters associated with the measurement. The angular data in the database are transformed from the detector coordinate system (or the local system) to the geographic coordinates (or the global system). Furthermore, the measurement log associated with the measurement is stored in the database. The log provides all operating parameters of the detector as a function of time, so its operation is easy to monitor. The muon tracking procedure with all its parameters and the background subtraction processes can also be monitored, and the data from the positioning and orientation of the detector can be obtained.

One kind of exceptional case is 1D muography (i.e., the total muon flux). In a very simplified view, it is also possible to construct density information about the very large rock volume by measuring simple muon fluxes (i.e., the sum of all initial muon directions as a function of time) in the different locations below the studied rock volume. In other words, the muon data lacks the angular data completely (which can thus simply be omitted) and consists of just four parameters: flux = flux(x, y, z). However, the number of detector positions can be large and as combined they can provide a 3D picture about the rock volume of interest. This simple method is put to the test in the AGEMERA project, and we are currently investigating the possibilities to extract meaningful density data using geometrically a very simple approach for mapping large rock volumes.

4.5.1 Litology related muographic data

Muography as a method relates to the geological structures, ore bodies, and such general physical characteristics of rocks that are investigated through the extracted density distributions (Beni et al., 2023). The object (or the target) can be delimited on the basis of detected density contrasts. Furthermore, the method can also be employed for the search of voids (Holma et al., 2022). The approximate density distribution is produced in voxelized form by software using Bayes' principle. The problem is a so-called inverse problem, i.e., the density voxels (cubes in 3D) are calculated using a set of observations that resulted in the observed muographic images. In other words, it is called an inverse problem because it is solved by starting with the effects which are used to determine the causes that produced them and that is why it is also called an inversion. During the inversion, the estimated variances of the density values are also derived.

The flowchart of muographic data processing together with different data types is shown in Figure 9. The data processing begins with the measured data and ends to transformation via inversion and classification. The inputs to the inversion can be muon counting rates, muon fluxes, rotated muon fluxes, density lengths and apparent densities.

In a broader and more detailed context the data flow in the muographic module can also be explained in terms of functions, data, database and outer (or external) databases. This is illustrated in Figure 10. One notes that the outer databases provide information that is mainly related to the data that are independent on the site, or one particular measurement and therefore those can be used as external sources of information, and the number of such databases can vary between different measurements depending on the number of available databases.

The used inversion algorithm relies on prior information from the current geological model (Bayesian principle). The density distributions produced in the inversion are categorised, again based on the assumed density of the target minerals and rocks (usually 2 – 10 different density categories). After appropriate coordinate transformations these categories are transferred to the main project database, which provides services (visualisation) to assist the geological interpretation. It is also worth noticing that, in addition to the major rock categories (and their different densities), the cavity and fracture systems associated with tectonics and karstification can also be explored by muography

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

and the results of the geological interpretation can also assist in the selection of further measurement points improving the final inversion results.

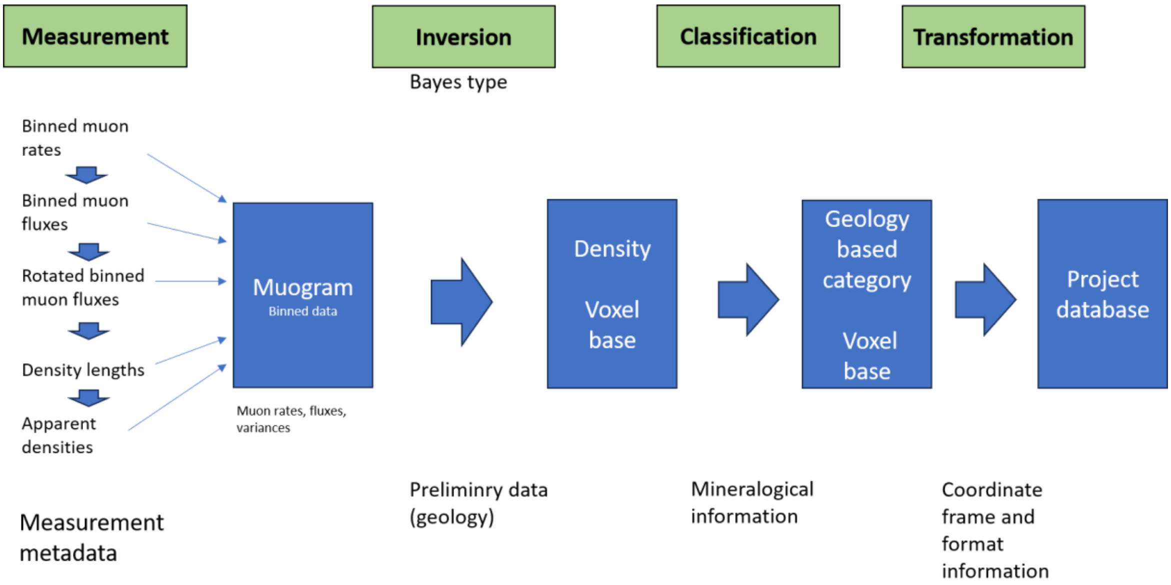
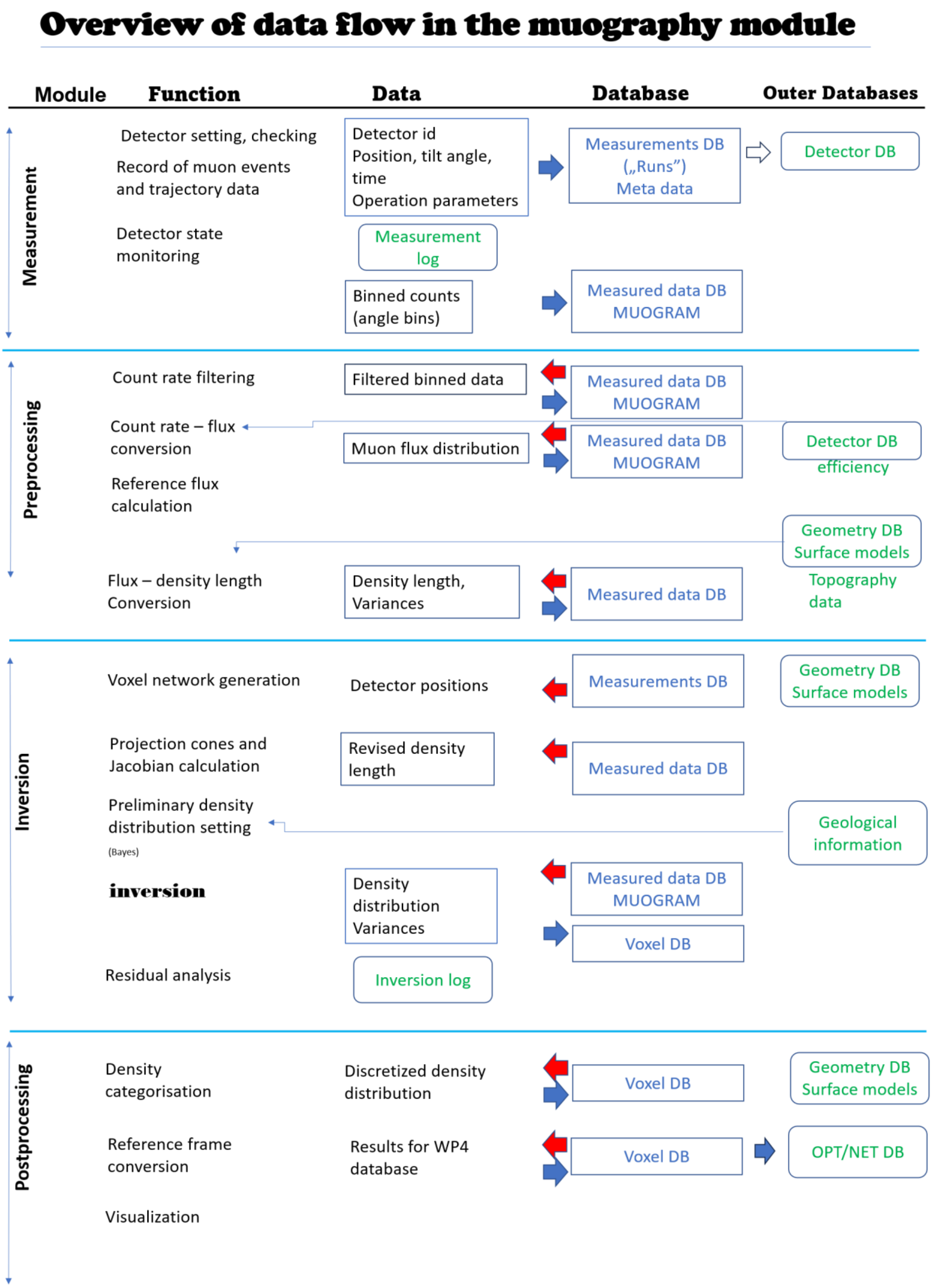


Figure 10. Flowchart of muographic data processing and data types

(Reference: Muon Solutions Oy)

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.



Overview of data flow in the muography module

Module	Function	Data	Database	Outer Databases
Measurement	Detector setting, checking Record of muon events and trajectory data	Detector id Position, tilt angle, time Operation parameters	Measurements DB („Runs”) Meta data	Detector DB
	Detector state monitoring	Measurement log		
		Binned counts (angle bins)	Measured data DB MUOGRAM	
Preprocessing	Count rate filtering	Filtered binned data	Measured data DB MUOGRAM	
	Count rate – flux conversion	Muon flux distribution	Measured data DB MUOGRAM	Detector DB efficiency
	Reference flux calculation			Geometry DB Surface models
	Flux – density length Conversion	Density length, Variances	Measured data DB	Topography data
Inversion	Voxel network generation	Detector positions	Measurements DB	Geometry DB Surface models
	Projection cones and Jacobian calculation	Revised density length	Measured data DB	
	Preliminary density distribution setting (Bayes)			Geological information
	inversion	Density distribution Variances	Measured data DB MUOGRAM Voxel DB	
	Residual analysis	Inversion log		
Postprocessing	Density categorisation	Discretized density distribution	Voxel DB	Geometry DB Surface models
	Reference frame conversion	Results for WP4 database	Voxel DB	OPT/NET DB
	Visualization			

Figure 11. Overview of data flow in the muographic module

(Reference: Muon Solutions Oy)

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

4.5.2 Data fusion

The integrated database contains the verified data relevant to the geological information in a properly structured format (see Data Integration chapter for more details). These data are:

- Map (2D) data (morphology)
- Previously known data on geological structure and rocks (prior information)
- Rock physics data from geophysical measurements
- Sampling and drilling data, petrological data

These data (sometimes not all these are available, are used as approximated, interpolated or extrapolated data, or are omitted for simplicity) should be weighted according to their reliability in the geological interpretation when building the geological model (the goal of fusion). The reliability of the model is mainly determined by the measurement related errors, the resolution of the method and the approximations of the interpretation (inversion) models.

In the case of geophysical data, fusion also means defining the relationship between geological data and rock physics data. These can be used (in combination) to delineate, for example, rock block formations, fault zones, weathered regions, etc., depending on the extent to which rock quality (composition, etc.) is clearly represented in the mapping to physical properties. The model building process is also determined by the resolution and investigation depth of the geophysical methods used to map the geological objects under investigation (Hall & McMullen, 2004).

The fusion of geophysical data is thus the construction of a rock model, a geological model, from a spatial model of overlaying rock physical properties distributions. The relationship between physical rock properties and geology can also be clarified through rock sampling and prior knowledge of the geological processes in the areas under investigation. In the process of fusion, it may be possible to separate the research of the structure (layers, seeds, etc.) and the processing of information on the rock quality.

It is also worth noticing that estimates based on petrophysical equations (direct approach or inversion), traditional multivariate statistics (clustering) and machine learning methods can also be used in the fusion process (the so-called joint interpretation of geophysical data) in order to improve the final interpretation.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

4.6 Data Types Provided by Satellites

Satellites provide a wide range of data types, which are essential for various Earth observation, scientific research, and practical applications. These data types are collected by various Earth-observing satellites, both government-operated and commercial, and are used by governments, research institutions, businesses, and individuals for a wide range of applications, including environmental monitoring, agriculture, urban planning, disaster management, and scientific research. Access to satellite data from open access missions has become increasingly accessible and essential for addressing global challenges and advancing our understanding of Earth and, from the aspect of this project, the topic of critical Raw Materials. Satellite data can be used in every step of the project lifecycle, from remote sensing exploration, to detailed studies of pre-selected regions of interest, to mineral and resource mapping, environmental monitoring during and post closure. Depending on the goal, satellite data types can be categorized in several ways, but the most important when discussing CRM are optical and radar imagery.

Other categories of satellite data also include Light Detection and Ranging (LiDAR) data sensors (measures the distance between the satellite and the Earth's surface), atmospheric data (temperature, humidity, and atmospheric composition for weather forecasting, climate monitoring, and air quality assessments), oceanographic data (data on sea surface temperature, sea level, ocean currents, and marine ecosystem health), navigation and positioning data, communication data (global connectivity for telecommunication, internet access, and broadcasting services) and emergency and disaster monitoring (monitoring natural disasters like hurricanes, earthquakes, and wildfires).

In addition to data types, satellites also offer various levels of data processing and image products, including raw data, orthorectified imagery, and derived products. Researchers, scientists, and organizations can access and utilize openly available data for a wide range of applications, making them valuable resources for Earth observation and analysis. Commercial data, on the other hand, can be sometimes needed for detailed analysis with extremely high spatial resolution (less than 4 m) in advanced stages of mineral prospecting, when potentially fruitful regions of interest have been identified by other methods.

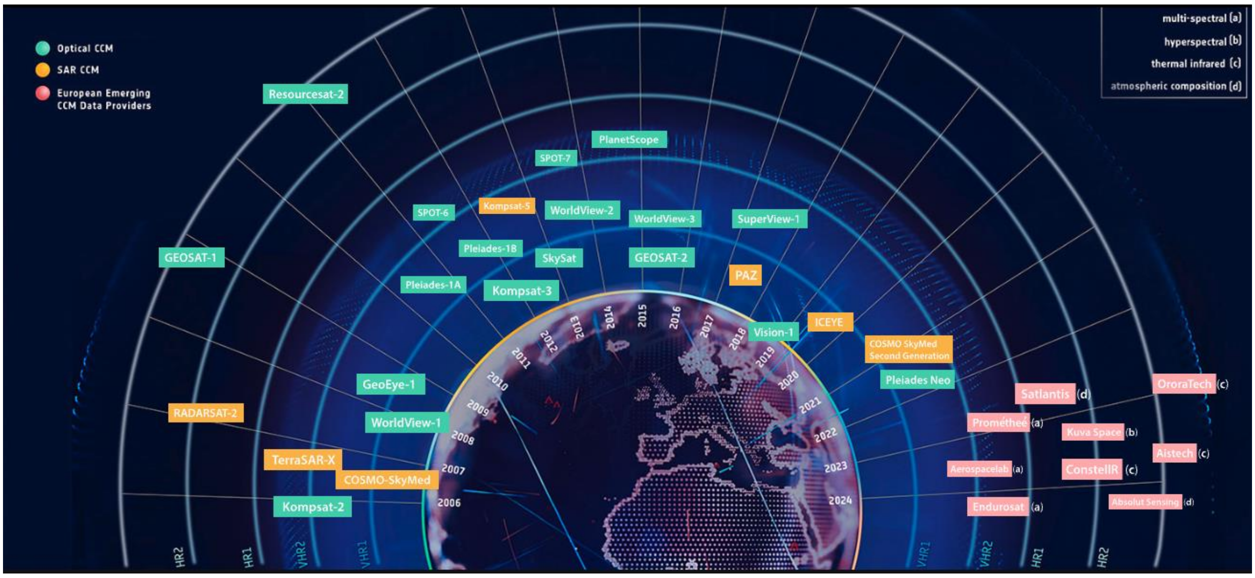


Figure 12. Copernicus Contributing Missions

https://spacedata.copernicus.eu/web/guest/contributing_missions

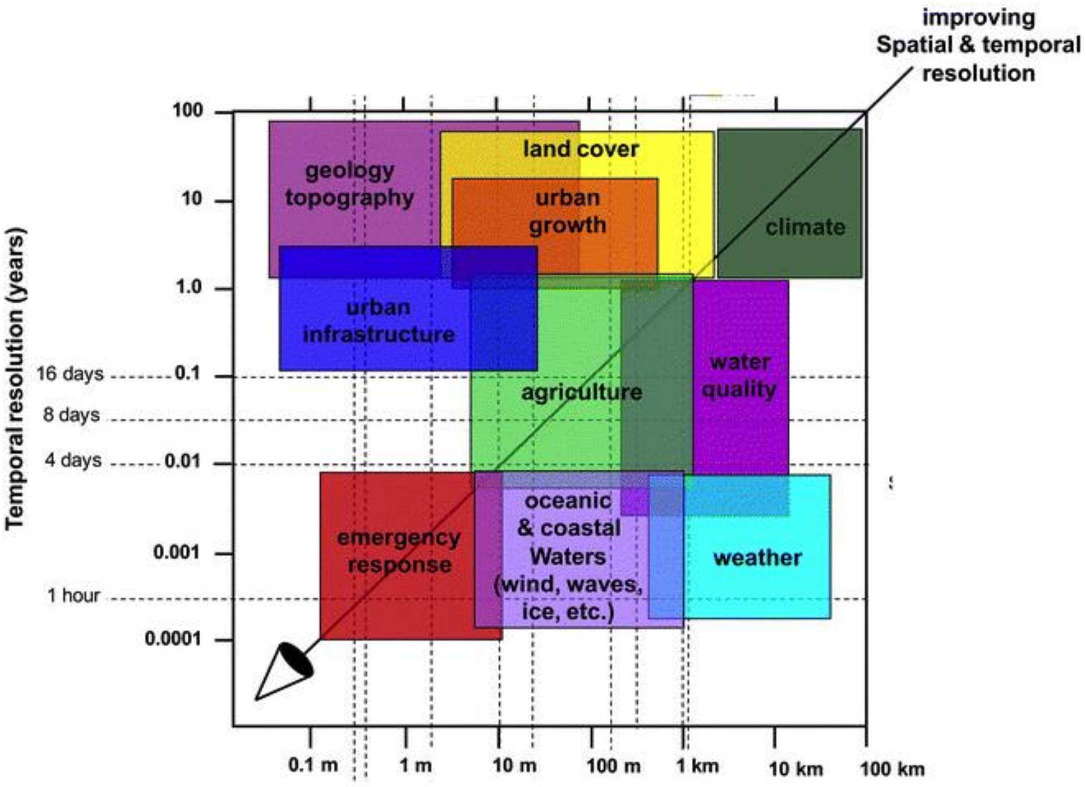


Figure 13. An overview of spectral, spatial and temporal resolution of satellite imagery

Adopted and modified from Kadhim, N., Mourshed, M. & Bray, M. (2016)

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

4.6.1 Optical Imagery

Optical satellite imagery refers to images captured by satellite (or UAV) with a passive optical sensor based on the reflection of sunlight from the Earth's surface. One of the key characteristics of optical imagery is the detection of light in the visible spectrum, even though it is not limited only to it:

- Visible and Near-Infrared (VNIR) Imagery captures light in the visible and near-infrared spectrum and is useful for tasks like land use and land cover classification, vegetation monitoring, and urban planning.
- Shortwave Infrared (SWIR) Imagery is employed for mineral exploration, moisture content estimation, and geological mapping.
- Thermal Infrared (TIR) Imagery measures the heat radiation emitted by objects on Earth's surface. It is used for applications like detecting hotspots in forest fires, monitoring urban heat islands, and estimating land surface temperature.

Many optical satellites have various capabilities regarding their spectral resolution:

- Multispectral Imagery consists of multiple narrow bands across the electromagnetic spectrum, providing valuable information for various applications.
- Hyperspectral Imagery offers an even larger number of narrow spectral bands, allowing for detailed analysis of materials and land features.
- Panchromatic Imagery, contrary to multispectral and hyperspectral, records data across a broad range of visible light wavelengths, resulting in a single-band high spatial resolution image, mostly used for detailed visual interpretation and analysis, such as urban planning or land use and land cover classification.

Another key feature of satellite optical imagery that should be emphasized is its weather dependency; more precisely dependency on clear weather with minimal cloud coverage and daylight, making its use limited to an extent. Combined with limited temporal resolution and specific geographic and climatological conditions of certain areas, for example rainforests, it may become increasingly difficult to obtain a useful optical satellite image.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

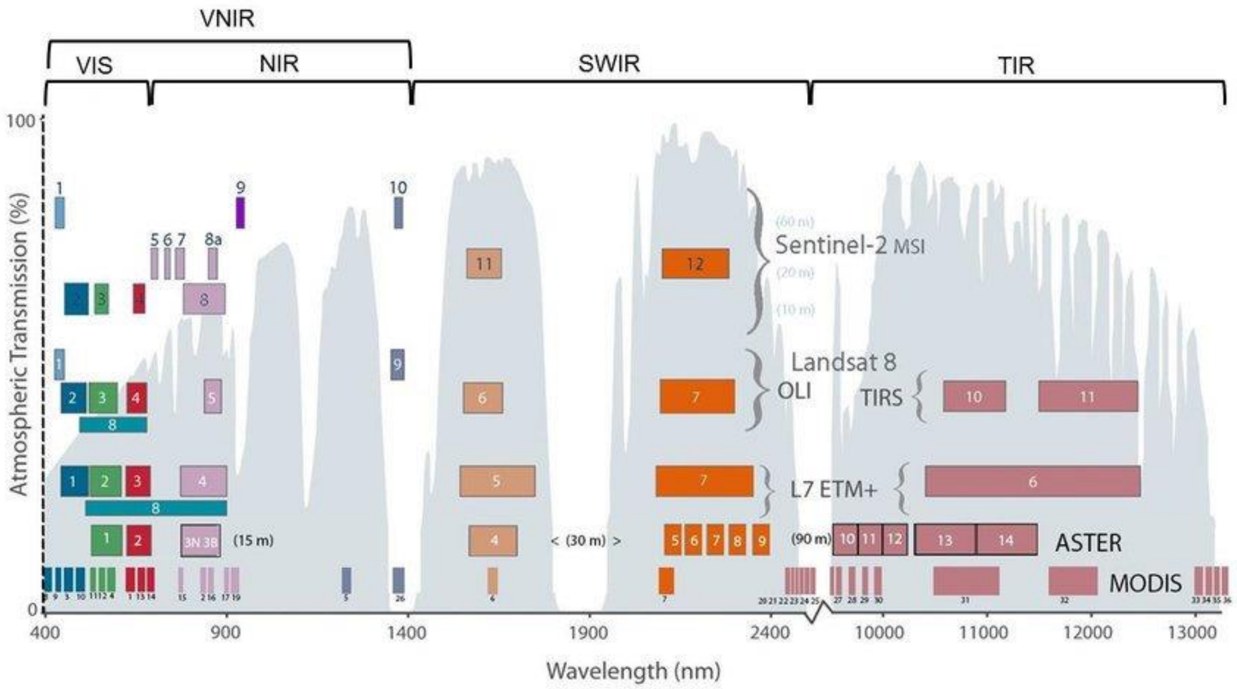


Figure 14. Spectral resolution of currently available optical satellite sensors grouped by different domains of the electro-magnetic spectrum (VIS = visible, NIR = near infrared, VNIR = visible near infrared, SWIR = shortwave infrared, TIR = thermal infrared)

(Friedl, P. 2020)

According to UCS records, as of December 2022, there are around 1030 active Earth Observation Satellites with different orbits, revisit times, using different instruments and sensors, belonging to different owners and countries. While there are more detailed lists of all (or most) of available satellite products in many publications, here are presented only the most common open-source optical satellites whose products are integrated with the AGEMERA platform backend component: Sentinel-2, ASTER, Landsat-8 and WorldView-3.

Sentinel-2 is a Copernicus Earth observation mission that provides multispectral optical imagery, including visible, near-infrared, and shortwave infrared bands, at high spatial resolution (10 - 60 m) over land and coastal waters. The constellation consists of two twin satellites, Sentinel-2A (S2A) and Sentinel-2B (S2B). The temporal resolution of each satellite is 10 days and 5 days when the acquisitions are combined. Sentinel-2 provides multispectral images with 13 bands in the visible, near infrared, and short-wave infrared part of the spectrum. It is commonly used for land cover classification, agriculture monitoring, and environmental assessments. Additional information on Sentinel-2 can be found at: https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/sentinel-2

ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) is a sensor onboard NASA's Terra satellite that flies in a sun-synchronous polar orbit. ASTER utilizes a unique combination of wide spectral coverage and high spatial resolution in three different modes:

- Visible and Near Infrared (VNIR) used for land cover classification, geological mapping, and vegetation analysis
- Shortwave Infrared (SWIR) useful for mineral identification, volcanic monitoring, and soil moisture estimation
- Thermal Infrared (TIR) is used to measure land surface temperature, which is valuable for various environmental and geological applications

ASTER products are distributed from the Land Processes Distributed Active Archive Center (LP DAAC) and are produced from on-demand data acquisition requests. Additional information on ASTER can be found at: https://asterweb.jpl.nasa.gov/

Landsat-8 is an American Earth Observation satellite launched as a collaboration between NASA and the United States Geological Survey (USGS) and it carries the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS).

- OLI provides data in multiple spectral bands, including visible, near-infrared, and shortwave infrared.
- TIRS measures thermal infrared radiation, allowing for the calculation of land surface temperature and heat energy flux.

Landsat-8 data is widely used for land cover classification, agriculture, forestry, and environmental monitoring. Additional information on Landsat-8 can be found at: https://www.usgs.gov/landsat-missions/landsat-8

WorldView-3 is the industry's first multi-payload, super-spectral, high-resolution commercial satellite with the daily revisit time. Near infrared and short-wave infrared can be used to identify the difference in structural features of the earth's surface. In addition, multispectral imaging and thematic mapping allow researchers to collect reflection data and absorption properties of soils, rock, and vegetation. This data could be utilized by trained photo geologists to interpret surface lithologies, identify clays, oxides, and soil types from satellite imagery. Additional information on WorldView-3 can be found at: https://earth.esa.int/eogateway/missions/worldview-3

4.6.2 SAR Imagery

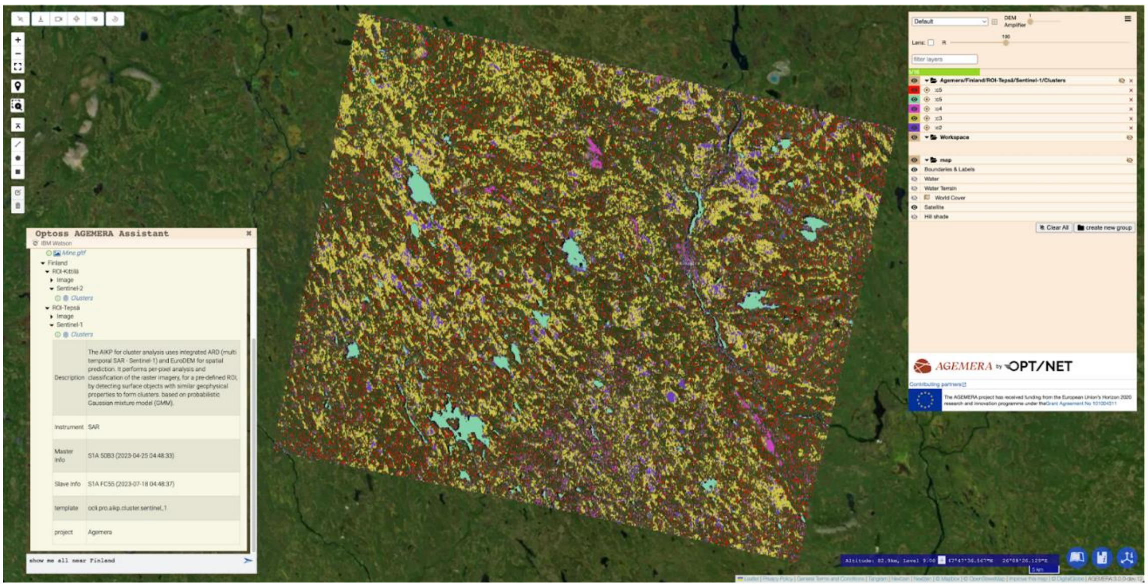
Synthetic Aperture Radar (SAR) Imagery is acquired by active radar sensors and doesn’t have a cloud/weather limitation of optical imagery. It is particularly useful for applications such as ground deformation monitoring, change detection, natural disaster monitoring or enhanced maritime surveillance. Its ability to provide consistent data in all weather conditions, make SAR imagery an indispensable tool for remote sensing and earth observation. Similarly, to satellites carrying optical instruments, there are currently many satellite missions that capture SAR imagery, but the most commonly used open-source is Sentinel-1, a part of Copernicus Earth observation mission.

Sentinel-1 is a radar satellite constellation made up of two satellites: Sentinel-1A (S1A) and Sentinel-1B (S1B) sharing the same orbital plane with a 180° orbital phasing difference, inclined 89° to the equator with a 12 day repeat cycle. It carries the Synthetic Aperture Antenna Radar (SAR) operating in different acquisition modes depending on the type of surface being mapped and on the acquisition schedule. The most common type over the land is IW Swath mode, consisting of three swaths and 9 bursts per swath. Additional information of Sentinel-1 can be found at: https://sentinel.esa.int/web/sentinel/missions/sentinel-1.

4.6.3 AIKPs integrated in the AGEMERA framework

Combining satellite data acquisition and image preprocessing with AI capabilities, AGEMERA platform offers several Earth Observation Products available in a form of AI Knowledge Packs. These AIKPs allow for the possibility of workflow automatization as every step from data acquisition through AI processing to data publishing on a web-based platform is incorporated into the process.

Unsupervised Clustering Analysis performs the spatial prediction from integrated pre-processed Analysis Ready Data (ARD) (either Level-1 Sentinel-1 SAR or Sentinel-2, ASTER, Landsat-8 optical imagery). In some more details, it is a per-pixel analysis and classification of the raster imagery, for a predefined region of interest, by detecting surface features with similar geophysical properties to form land class clusters. The clustering is based either on a probabilistic Gaussian mixture model or Bayesian Gaussian Mixture algorithm. The clustering map is more helpful than single-attribute geological maps for geological mapping and interpretation. Each cluster is characterized by multigeophysical properties and can be associated with certain geological attributes or rock types based on existing geological maps, field data and rock sample analysis.



Optoss AGEMERA Assistant

IBM Watson

- Mine.gtif
- Finland
  - ROI-Kittilä
    - Image
    - Sentinel-2
      - Clusters
  - ROI-Tepsä
    - Image
    - Sentinel-1
      - Clusters

Description	The AIKP for cluster analysis uses integrated ARD (multi-temporal SAR - Sentinel-1) and EuroDEM for spatial prediction. It performs per-pixel analysis and classification of the raster imagery, for a pre-defined ROI, by detecting surface objects with similar geophysical properties to form clusters, based on probabilistic Gaussian mixture model (GMM).
Instrument	SAR
Master info	S1A 50B3 (2023-04-25 04:48:33)
Slave info	S1A FC55 (2023-07-18 04:48:37)
template	ocli.pro.aikp.cluster.sentinel_1
project	Agemera

show me all near Finland

Figure 15. Unsupervised Machine Learning clustering analysis based on Sentinel-1 SAR imagery

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

Credits: OPT/NET B.V. (2023)

Composites are RGB composite images composed by assigning different bands (or band ratios) to RGB channels from a single Sentinel-1 or Sentinel-2 image. Currently, for Sentinel-1, the only available combination is Polarimetric Synthetic Aperture Radar (PolSAR), an advanced imaging radar system that provides scattering information under different combinations of wave polarizations. For Sentinel-2, the possible band combinations are true colour, false colour, SWIR, agriculture, geology, bathymetry, atmospheric penetration, healthy vegetation, land/water and snow, each emphasizing different aspects and characteristics of a selected region of interest.

Scalar index produces single-band rasters based on predefined channel arithmetic expressions. These speeds up the process of creating scalar index images manually dramatically, reduces the possibility of error and allows, in the future, for automatization of the process. Of course, the predefined band arithmetic expressions are defined by the available bands, so the AIKP is, for now, available only for ASTER and Sentinel-1 EOD. For example, band arithmetic expressions available for ASTER imagery that can be extremely useful in mineral prospecting are quartz, carbonate, ferric, silica, mafic or ultramafic indices.

PCA is a dimensionality reduction method that is often used to reduce the dimensionality of large data sets, trading a little accuracy for simplicity. In case of ASTER optical imagery, for example, it utilizes selective principal components analysis (SPCA), in which the standard PCA is applied in a selective way using only bands that theoretically describe a mineral feature and avoiding bands that would cause interference such as vegetation. To that end, four bands (two for high reflectance, two for low reflectance) are used as inputs to detect sericite, chlorite, and epidote.

Soil moisture index and time series (oRVI) are based on Sentinel-1 imagery and can be used to examine the internal structure of the soil moisture based on multi-temporal SAR data for the selected areas of interest. oRVI is defined as a ratio of cross-polarization to total power from all polarization channels. As a ratio, RVI has less sensitivity to radar measurement geometry and topography and remains insensitive to absolute calibration error in radar data.

Additional data from other sources, such as geochemical analysis and geophysical exploration can also be ingested into the platform, through AIKPs specifically designed for DEM, single-band and multiband raster publishing in georeferenced GeoTIFF file format.

5. Conclusions

5.1 Summary

This report, designated as Deliverable 4.2 (D4.2) of the AGEMERA project, focuses on explaining the methodologies and outcomes of Task 4.2 (T4.2) Data Fusion. The essence of T4.2 lies in its integrative approach, aiming to unite diverse and heterogeneous datasets from Work Packages (WPs) 1 and 3 to foster an optimal understanding of the geological, geochemical and geophysical features within each field trial area. This integration process is closely related the work of T4.1, which concentrates on the data processing. T4.2 expands this scope by incorporating additional geological and geophysical information from WPs 1 and 3, respectively, alongside satellite-based spatial data.

This report presents a comprehensive and integrated workflow detailing the amalgamation process of geological and geophysical datasets within the AGEMERA project through data fusion methodologies. It also ensures that the workflow is adaptable and hence beneficial for similar data fusion tasks beyond the confines of the current project.

5.2 Implications for Future Research

Future research can be built upon the foundation laid by the data fusion methodologies employed in this project. The adaptability of the workflow presented herein suggests its applicability to a wide array of geological contexts, extending beyond the specific field trial areas of the AGEMERA project. Researchers are encouraged to explore the integration of additional data types and sources, potentially including emerging technologies and novel geophysical methods, to further refine and expand the capabilities of data fusion in geological research.

Moreover, the linkage established between computer-based modelling (T1.4) and the development of new and improved mineral system models for CRM ore deposit types (T1.5) highlights the importance of interdisciplinary collaboration. Future research initiatives should continue to foster synergies between different scientific disciplines, leveraging the strengths of each to achieve a more holistic understanding of mineralisation processes and other geological phenomena.

The implications of this report also extend to the realm of resource management and sustainability. By enhancing the precision and depth of geological models, the methodologies discussed herein can contribute to more efficient and environmentally conscious exploration practices. This is particularly valid in the context of CRMs, which are essential for the advancement of green technologies and sustainable development. It is crucial to acknowledge that if the exploration methodologies applied for CRM exploration are not suboptimal in terms of carbon footprint and other environmental and community-related aspects, it could weaken the argument that exploration of these resources is important. Therefore, the adoption of sustainable and responsible exploration practices is not only beneficial for the environment and communities but also vital for reinforcing the significance of CRMs in supporting sustainable development goals.

5.3 Recommendations

The following recommendations are proposed to enhance the effectiveness of future research and exploration activities of CRMs and other raw materials:

1. **Adoption of integrated data fusion approaches:** This approach has proven effective in providing a comprehensive understanding of geological features and should be considered a best practice in the field.
2. **Sustainable exploration practices:** It is imperative to prioritise sustainable and environmentally conscious exploration practices. Future research should focus on developing and implementing methodologies that minimise the carbon footprint and ensure minimal disruption to local communities and ecosystems.
3. **Investment in innovative technologies:** Continued investment in innovative technologies, such as drone geophysics, cosmic-ray muography and passive seismics, is recommended. These technologies have shown promise in enhancing the precision of geological models and should be further explored and developed.
4. **Cross-disciplinary collaboration:** Encourage and facilitate cross-disciplinary collaboration among geologists, geochemists, geophysicists, and other relevant experts. Such collaboration can lead to more robust and holistic models, ultimately improving the accuracy and efficiency of CRM exploration.

By following these recommendations, future research and exploration activities can not only advance the understanding and extraction of CRMs but also ensure that such endeavours are conducted in a manner that is responsible, sustainable, and beneficial to all stakeholders involved.

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

References

Balázs L, Nyitrai G., Surányi G., Hamar G., Barnaföldi G., Varga D., (2023) 3D Muographic Inversion in the Exploration of Cavities and Low-density Fractured Zones, arXiv:2309.12057, https://doi.org/10.48550/arXiv.2309.12057

Beni, T., Borselli, D., Bonechi, L. et al. (2023) Transmission-Based Muography for Ore Bodies Prospecting: A Case Study from a Skarn Complex in Italy. Nat Resour Res 32, 1529–1547 (2023). https://doi.org/10.1007/s11053-023-10201-8

Berbellini, A., Schimmel, M., Ferreira, A.M.G., Morelli, A. (2019) Constraining S-wave velocity using Rayleigh wave ellipticity from polarization analysis of seismic noise. Geophysical Journal International, 216(3), 1817-1830, 10.1093/gji/ggy512

Castanedo, F. (2013). A Review of Data Fusion Techniques. The Scientific World Journal, 2013, 19. doi:10.1155/2013/704504

Friedl, P. Derivation of Glaciological Parameters from Time Series of Multi-Mission Remote Sensing Data—Applications to Glaciers in Antarctica and the Karakoram. Ph.D. Thesis, Friedrich-Alexander-University of Erlangen-Nürnberg, Erlangen, Germany, 2020

Hall, D. L. & McMullen, S. A. H. (2004). Mathematical Techniques in Multisensor Data Fusion, 2nd ed., Artech House, Norwood, Massachusetts

Holma M. (2023a) Beyond Traditional Methods: Muography's Integration into Applied Geophysics for Enhanced Mineral Discovery, 2023, Conference: Sovelletun geofysiikan XXIV neuvottelupäivät, At: 22.11.2023, Oulu, Vuorimiesyhdistys, pp. 24-27

Holma M. (2023b) Applications of cosmic-ray muon imaging in Earth Sciences, March 2023, Conference: GeoDays 2023, At: 14th-17th March 2023, Helsinki, Finland

Holma M., Zhang Z., Kuusiniemi P., Loo K., Enqvist T. (2022), Future Prospects of Muography for Geological Research and Geotechnical and Mining Engineering, Book Editor(s): L Oláh, Hiroyuki K. M. Tanaka, D Varga, https://doi.org/10.1002/9781119722748.ch15

Jones, G.A., Ferreira, A.M.G., Kulessa, B., Schimmel, M., Berbellini, A., Morelli, A. (2021). Uppermost crustal structure regulates the flow of the Greenland Ice Sheet. Nature Communications, 12, 7307. 10.1038/s41467-021-27537-5

Kadhim, N., Mourshed, M. & Bray, M. (2016). Advances in remote sensing applications for urban sustainability. Euro-Mediterranean Journal for Environmental Integration. 10.1007/s41207-016-0007-4

Lesparre N., Gibert D., Marteau J., Déclais Y., Carbone D., Galichet E., (2010) Geophysical muon imaging: feasibility and limits, Geophysical Journal International, Volume 183, Issue 3, December 2010, Pages 1348–1361, https://doi.org/10.1111/j.1365-246X.2010.04790.x

Nuñez, E., Schimmel, M., Stich, D., Iglesias, A. (2020) Crustal velocity anomalies in Costa Rica from ambient noise tomography. Pure and Applied Geophysics, 177, 941-960, 10.1007/s00024-019-02315-z

Romero, P., Schimmel, M., (2018). Mapping the basement of the Ebro Basin in Spain with seismic ambient noise autocorrelations. Journal of Geophysical Research, 123, 5052-5067, 10.1029/2018JB015498

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

Schimmel, M., Stutzmann, E., Gallart, J. (2011). Using instantaneous phase coherence for signal extraction from ambient noise data at a local to a global scale, Geophysical Journal International, 184, 494-506. 10.1111/j.1365-246X.2010.04861

The National Land Survey of Finland (NLS) https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/expert-users/product-descriptions/laser-scanning-data-05-p, accessed 20 Oct. 2023.

Varga D., Nyitrai G., Hamar G., Galgóczi G., Oláh L., Tanaka H.K.M., Ohminato T. (2020) Detector developments for high performance Muography applications, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 958, 2020, 162236, ISSN 0168-9002

Link referenced to thought out the text that offer more information about the subject

SAFE file format
HDF file format
GeoTIFF file format
NetCDF file format
World Geodetic System
EPSG:4326
UTM coordinate system
Converting UTM zones to EPSG codes
EPSG:4979
EPSG:4978
EPSG:5773
EPSG:5798
EPSG:3855
Copernicus Contributing Missions
Sentinel-2
ASTER
Landsat-8
WorldView-3
Sentinel-1

This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.