AGEMERA

Critical Raw Materials for
a Resilient Europe

# DATA MANAGEMENT PLAN

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:** | D7.3.1 Data Management Plan |
| **Due date of deliverable:** | 31.01.2023 |
| **Actual date of submission:** | 03.02.2023 |
| **Deliverable Lead Partner:** | **University of Oulu** |
| **Dissemination level:** | Public |

## Author list

| Name | Organization |
|---|---|
| Jari Joutsenvaara | University of Oulu, Kerttu Saalasti Institute |
| Eija-Riitta Niinikoski | University of Oulu, Kerttu Saalasti Institute |

**Document History**

| Version | Date | Note | Revised by |
|---|---|---|---|
| 01 | 11.01.2023 | preliminary draft | JJ |
| 02 | 24.1.2023 | draft version for commenting | JJ |
| 03 | 31.1.2023 | Final version for commenting | JJ |
| 04 | 3.2.2023 | Final version | JJ |

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

# 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.

Publication information

The dissemination level of this document is PUBLIC.

# Executive Summary

The AGEMERA project implements a combination of multisource geoscientific data and data fusion and processing powered by data analytics, data fusion and machine-learning algorithms. The data is collected and gathered into the AGEMERA platform, which combines the public, e.g., earth observation data, project-wise collected new geoscientific field data, drone, muographic and ambient seismic tomography data and converts the information into actionable intelligence for mineral potential and mineral exploration data. The project surveys local communities' concerns and hopes regarding mineral exploration and compiles the generalised and anonymised results into an open-access database.

This is the first version of the data management plan and reflects the current view of the equipment and the data types needed for the successful execution of the project. This document is updated throughout the project. New equipment or data sources are introduced during the project to provide data to the repositories. However, the data types of those possible new instruments and data sources will most likely follow the types already listed in this document.

For intraproject workspace, the project uses Microsoft Teams, and Centre for Scientific Computing (CSC) cloud services are used for intermediate storage of big data sets. For ready data sets, the Zenodo repository is used for storing. In Zenodo, the public data sets will be open, and those with IPR claims, e.g., patenting-related, will have restricted access. The different data sets will be described in more detail in the Zenodo repository. Further re-use of the generated data will be established by providing the created open data through the well-established EGDI platform, a dedicated platform created and maintained by the EuroGeoSurveys.

# Table of Contents

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

# List of Acronyms

| | |
|---|---|
| AI | Artificial Intelligence |
| CRM | Critical Raw Material |
| CSC | Finnish Centre for Scientific Computing |
| D | Deliverable |
| DMP | Data Management Plan |
| EGDI | European Geological Data Infrastructure |
| JRC | European Commission´s Joint Research Centre |
| ML | Machine learning |
| RMIS | Raw Material Information System |
| Zenodo | A data platform for sharing project information |

# 1 Data summary

This is the first version of the data management plan and reflects the current view of the equipment and the data types needed for the successful execution of the project. This document is updated throughout the project. It may be that during the project, new equipment or data sources are introduced to provide data to the platforms (internal AGEMERA platform, Zenodo repository) and, for the final open products, the EGDI platform. However, the data types of those possible new instruments and data sources will most likely follow the types already listed in this document.

## 1.1 The data

### Purpose of the data

Agemera project uses innovative methods and technologies to unlock the EU's resource potential, improve public knowledge of critical raw materials, survey the local communities' concerns and hopes related to mineral exploration and mining, and promote environmentally friendly mineral exploration. The project develops new genetic minerals models to unlock the critical raw material potential within the six target countries.

The project generates and re-uses a combination of multisource geoscientific data, new generated data from new innovative exploration methods, field studies, and online and local surveys on social sustainability and responsibility. (see Fig. 1. For AGEMERA data process flow). See attachment 1 for the data collection sheet.

The AGEMERA platform combines the public, e.g., earth observation data, project-wise collected new geoscientific field data, drone, muographic and ambient seismic tomography data, into actionable intelligence for mineral potential and mineral exploration data. The AGEMERA platform mobilises the new datasets generated by in-situ measurements via rapid generation of integrated spatiotemporal Analysis Ready Datacubes (ARDs) (see Fig. 2.) and subsequent data fusion by on-demand mix-and-match with heterogeneous, resampled, and precisely co-registered sources from the other spatially relevant remote sensing sources. Data analytic methods and machine-learning algorithms are used to process the data and transfer the information into an easily understandable form. The platform is IPR protected.

The above data sets and possibly new ones will be described in more detail in the Zenodo repository (https://zenodo.org/communities/agemera/). Further re-use of the created data will be established by providing the created open data through the well-known EGDI platform, a dedicated platform created and maintained by the EuroGeoSurveys (https://www.europe-geology.eu/).

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

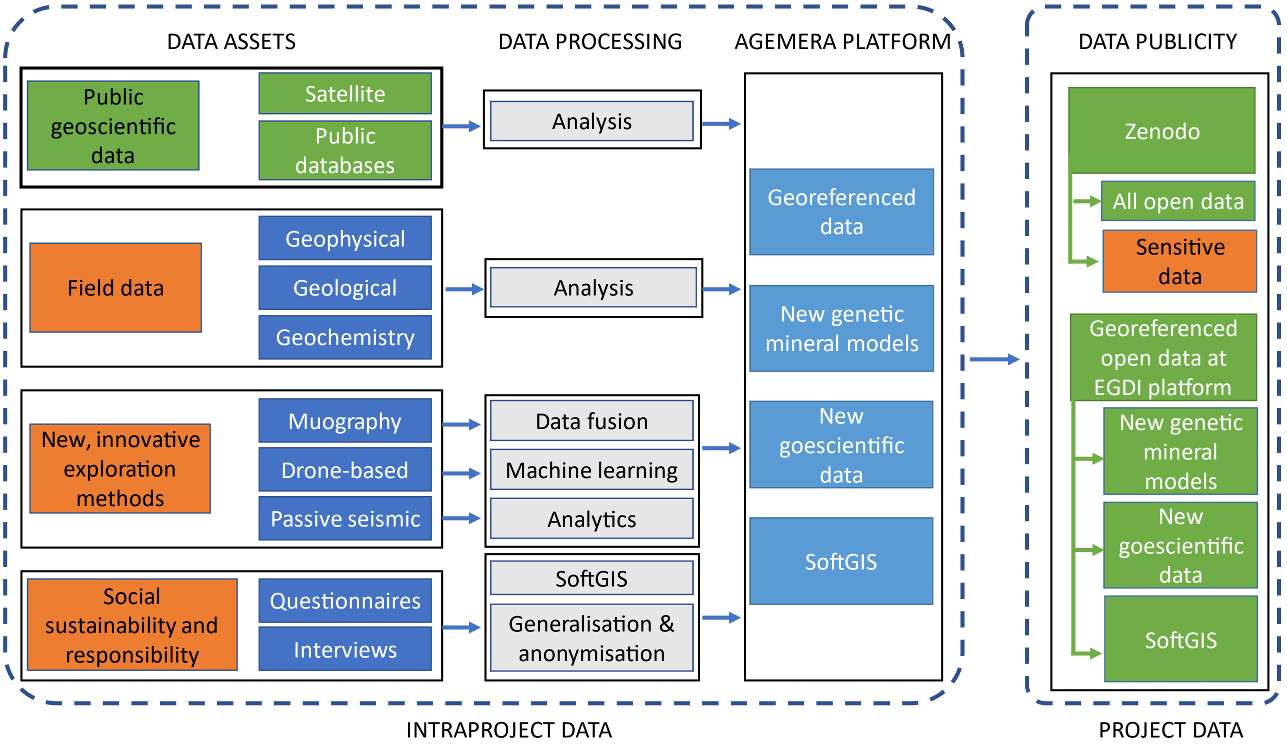


*Figure 1. The primary data sources for the AGEMERA project. The colours define the data publicity: green = public, field data is collected project-wise, and public results will be published. Technology developers' data is sensitive together with social data. Processed, public data will be published on the EGDI platform.*

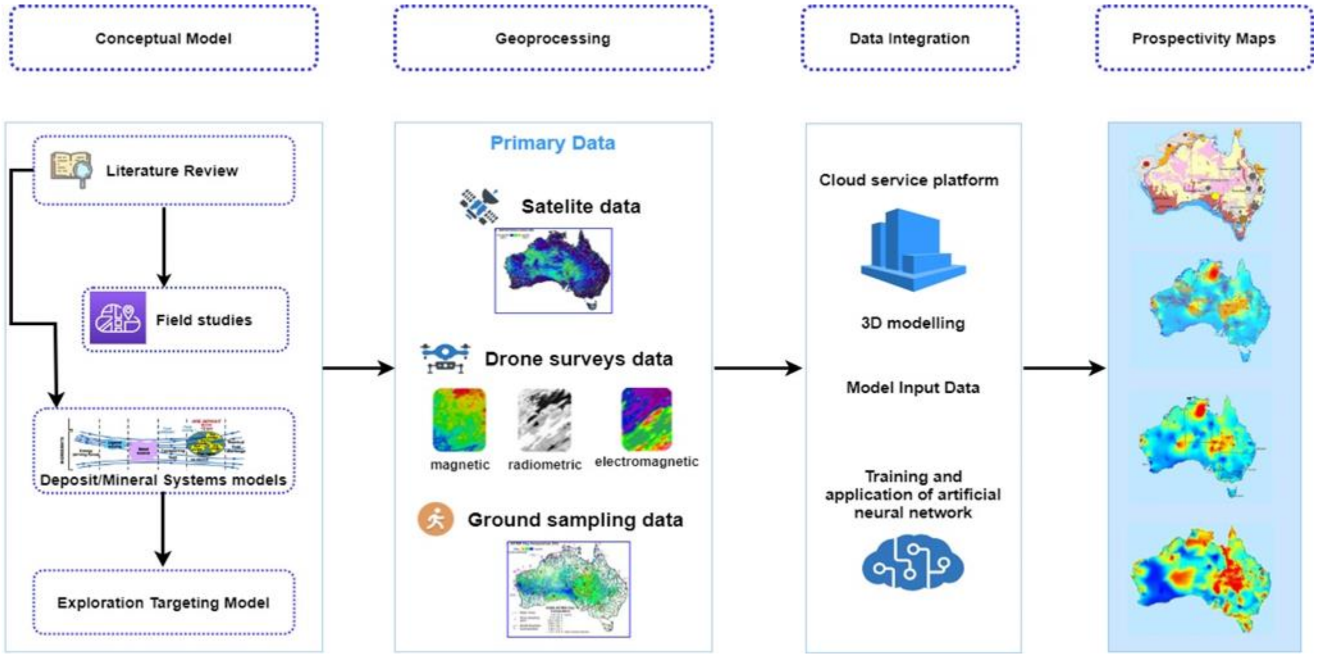


*Figure 2. Multi-dimensional datasets (hypercubes) from geophysical surveys.*

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

## 1.2 The data generation, formats and data usability

### Origin of the data

The project data may originate from open databases, trial site infrastructures, or mine infrastructures. We will ensure that we sign formal open data-sharing agreements with mines for those datasets we would like to place in open repositories.

### Data size

The estimated data storage requirement for the whole project data is in the order of Terabytes. The passive seismic data sets are in the Terabyte range. Drone data in the scale of hundreds of Gigabytes to a few Terabytes. Muographic data sets are in the order of 100s Megabytes. Public data sets to be used, e.g., Sentinel 1 & 2, public geoscientific depository data sets will be in the order of Gigabytes. Social sustainability and responsibility data are expected to be 10s to 100 Megabytes per target area.

### Data formats

Different data sources and instruments produce various types of data, mostly georeferenced. The project generates or re-uses satellite, geological, geophysical, geochemical, seismic, drone and muographic data.

Data formats include ASCII and comma-separated values (CSV) and GeoTIFF formats. However, CSV files do require separate metadata in the form of JSON files to be accompanied. GeoTIFF format allows storing the georeferenced metadata into the TIFF image´s metadata. TIFF stands for Tagged Image File Format and is a commonly used raster graphic file type. The digital elevation models (DEMs) are an example of information stored in GeoTIFF files. The GeoTIFF associates the evaluation (z) for every coordinate (x,y), and it also includes info on the used Coordinate Reference System (CRS). This allows the representation of pixel coordinates in real-world coordinates. For drone-based sensors, and in the case of AGEMERA, the electromagnetic, magnetic and radiometric maps are also in GeoTIFF formats.

### Data usability

We expect that our open datasets will be useful to

1) researchers (to progress the state-of-the-art further),
2) industrial partners  (to improve the sustainability of the exploration and mining processes and to improve discovery rates),  and
3) public authorities, local communities, and NGOs  (to support a  social license to operate).

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

# 2 FAIR data

## 2.1 Making data findable, including provisions for metadata

### Persistent identifier

Open data, results, and publications produced by the AGEMERA project will be made findable by adding file type-specific metadata using persistent Uniform Resource Locators (URL) and by assigning a Digital Object identifier (DOI). As the open materials in the project will be deposited in the Open Access repository Zenodo, the DOIs are automatically assigned.

### Keywords

In addition to the metadata, all open materials and results deposited in the repository will have associated search keywords. When possible, the keywords are chosen from datatype-specific vocabularies allowing machine-based harvesting and indexing.

### Metadata

The metadata will be done according to the templates of Zenodo. Muography metadata does not have a standardised format at the time of the creation of this version of Data Management. The metadata standard will be proposed during the project.

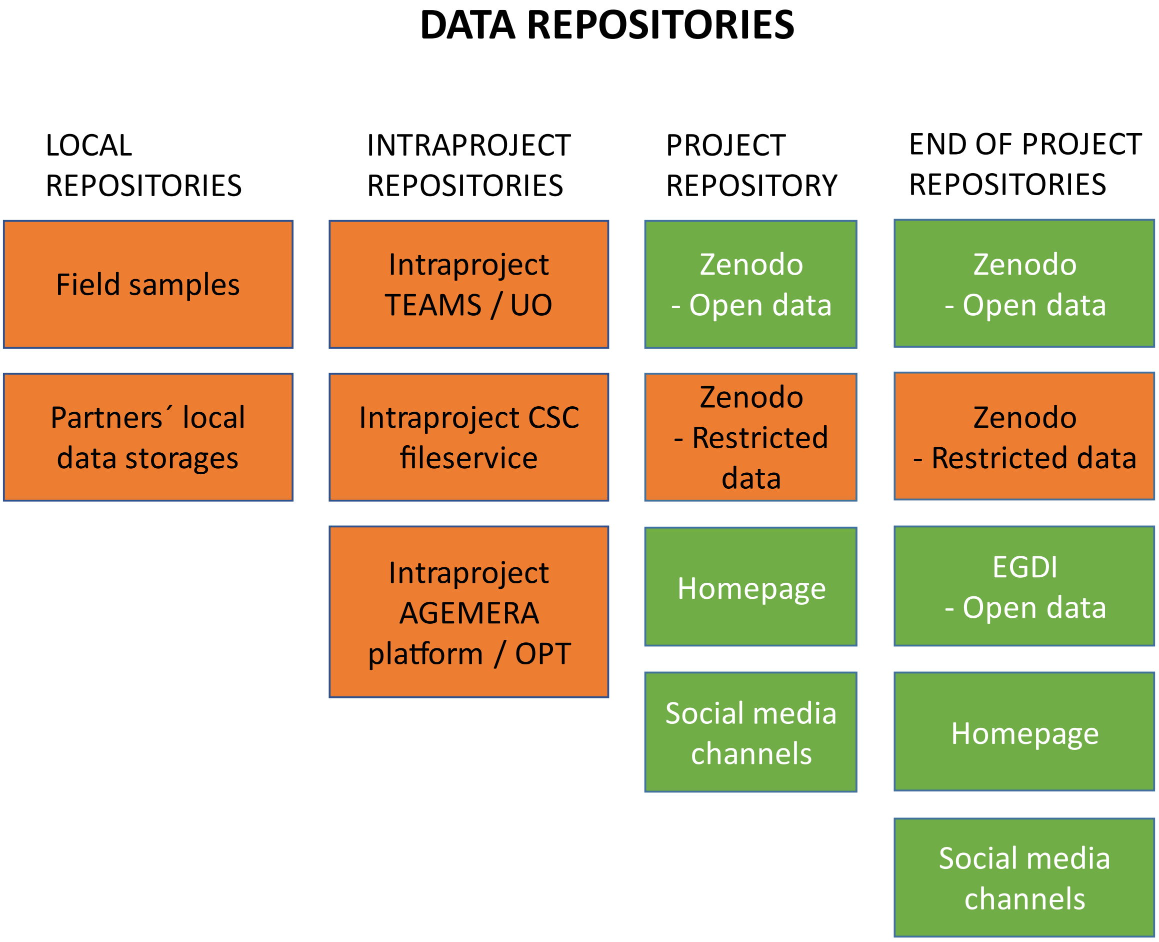
## 2.2 Making data accessible

### Repository

The project uses both intraproject and project repositories for data deposition and sharing. See Figures 1 & 3 for how the data will be deposited and shared. For intraproject data sharing, we use the Microsoft Teams online working environment. CSC cloud storage acquired for the project, and partners' local repositories are used for massive intermediate data sets. The AGEMERA platform is developed within the project to combine different data sets into actionable intelligence. The platform itself is IPR-protected due to future commercialisation.

Open GIS-based datasets and relevant geoscientific data sets will also be published on the Zenodo platform during the project. Zenodo accepts all kinds of research output, including publications, posters, presentations, software, media flashes and interactive materials such as lessons. At the end of the project, final open data sets are deposited on the EGDI platform with appropriate metadata. The EGDI platform is widely used by European projects to share their public geoscientific data sets. Using the platform ensures the reach of experts without the challenge of marketing and finding users for a new one. If seen necessary to deposit material elsewhere, e.g., institutional repositories, home pages etc., the data will have an identifiable URL.

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



DATA REPOSITORIES

| LOCAL REPOSITORIES | INTRAPROJECT REPOSITORIES | PROJECT REPOSITORY | END OF PROJECT REPOSITORIES |
|---|---|---|---|
| 🟧 Field samples | 🟧 Intraproject TEAMS / UO | 🟩 Zenodo - Open data | 🟩 Zenodo - Open data |
| 🟧 Partners´ local data storages | 🟧 Intraproject CSC fileservice | 🟧 Zenodo - Restricted data | 🟧 Zenodo - Restricted data |
|  | 🟧 Intraproject AGEMERA platform / OPT | 🟩 Homepage | 🟩 EGDI - Open data |
|  |  | 🟩 Social media channels | 🟩 Homepage |
|  |  |  | 🟩 Social media channels |

*Figure 3. The AGEMERA uses various repositories throughout the project. Local repositories are used to store generated data, and intraproject repositories are used to share data among the project consortium. Project repositories are used to store final datasets and promote the project and project outcomes. End of the Project repositories are used to give access to the project´s public data. Orange labelled data has restricted access due to sensitive data, and green labelled is open data.*

## Data

Each beneficiary must disseminate its results, unless it goes against their legitimate interests, by disclosing them to the public by appropriate means, including scientific publications (in any medium). However, this does not change the obligation to protect the results, confidentiality and security obligations, or obligations to protect personal data. Publication of data is formally agreed upon via an email to data management responsible from an authorised person from each partner. The list of authorised persons is stored in the project's shared folder in Teams.

All open data and deliverables marked 'public' in the Grant Agreement will be made public and published at Zenodo (after data is created) and at the end at the EGDI platform to ensure the comprehensive and versatile re-use potential of the created open data sets. Data sets with Intellectual Property Rights, e.g., leading to a patent, are closed and opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. In such cases, Zenodo allows data to be deposited under embargo status so that the repository restricts access until the end of the embargo period. As Zenodo offers data to be archived using a restricted access method, there is no need for dedicated access committee for the project.

Accessing the Zenodo repository and AGEMERA community page happens through https://zenodo.org/communities/agemera/. Access requires a web browser. Open data and publications can be accessed without logging in to the repository. Uploading files and accessing the restricted content requires login to the repository, and permissions can be given to selected users. Metadata is harvestable using the OAI-PMH protocol, and open APIs can be used to access data.

Access to intraproject repositories is controlled by the project coordinator and is given upon request. The Microsoft Teams online workspace and the CSC repositories require a login procedure. The coordinator can limit the accessibility to a single or limited number of users per working folder(s). The AGEMERA platform user access is controlled by the platform developer, who defines the access level to specific data sets or trial areas. The platform and the data within are not public.

### Metadata

Metadata will be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement. Most accompanying files are ASCII (text), CSV, GeoTIFF, or JSON data types. Free viewers are available in vast numbers.

## 2.3 Making data interoperable

The open data to be shared is interoperable by preferably having file types in either ASCII (text), CSV, GeoTIFF, or JSON file form and can be opened with free viewers or with open-source programs like QGIS, Python, R or even Notepad and equivalent text editors. There are no restrictions on the use of open published data. However, users are required to acknowledge the Consortium and the source of data in the possible resulting publications. Common data vocabularies, standards and methodologies are to be used to ensure the data's interoperability. If it is necessary to generate project-specific ontologies or vocabularies, mappings to more commonly used ontologies are provided.

## 2.4 Increase data re-use

The final public data sets will be made available on the EGDI platform, which already hosts EU-project-produced data sets from various projects. The platform is widely used to share public datasets, data descriptions, supporting documents, and open-access publications. As the end results will be made available through the EGDI platform, the data will remain accessible to third parties even after the project's end. Open-access publications and public deliverables will be available through Zenodo,

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

Excerpt from the https://www.europe-geology.eu/about-egdi/

> *“EGDI is EuroGeoSurveys’ (EGS) European Geological Data Infrastructure and is a central component of EGS’ strategy.*
>
> *EGDI provides access to Pan-European and national geological datasets and services from the Geological Survey Organisations of Europe. Through EGDI data from a number of European data harmonisation projects are accessible. EGDI was launched in June 2016 in a Version 1 and has since then been extended to include more data sets.*
>
> *The operation and maintenance of EGDI has in recent years been funded by EuroGeoSurveys and the operations, maintenance, and developments are carried out by the following EuroGeoSurveys members: GEUS, CGS, GeoZS, IGME, BRGM and BGS. The work in carried out in close cooperation with EGS’ Spatial Information Expert Group.*
>
> *EGDI formed the basis for the Information Platform developed in the GIP-project under the GeoERA programme which ran from July 2018 till October 2021. The GIP-project substantially extended the functionality of EGDI.*
>
> *In September 2022 a new 5 year Horizon Europe Coordination and Support Action called “Geological Service for Europe (GSEU)” started and this project will further develop EGDI.*
>
> *Data about raw materials from EGDI are accessible on the EC Joint Research Center’s Raw Materials System (RMIS). EGDI will furthermore provide the gateway from the Geological Survey Organisations and their geological data and digital services to the European Plate Observing System (EPOS) when that platform becomes operational.”*

The project will have mechanisms for quality assurance, which are described in the project handbook (D7.1.). The data quality will be validated by detailed characterisation of the data gathering setups.

# 3 Other research outputs

The geoscientific field data can also include taking samples for further analysis. An Excel sheet for the collected field data and samples is prepared to follow the data and sample storage within and after the project. See attachment 2 for the sample collection and storage sheet.

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

# 4 Allocation of resources

During the project, the raw data is stored on the local servers of the partners. Each of the partners carries their own local costs. We use CSC (centre for scientific computing, Finland) as an intermediate storage for large data sets needing storage or processing. The CSC service is free-of-charge through the coordinator University of Oulu, Finland. The AGEMERA platform´s cloud storage costs are covered by a project partner. The final products, open-access papers and pre-prints will be stored on the Zenodo platform. The Zenodo is free of charge and continues to host the materials even after the project. This is true for both open and sensitive materials. The final GIS data will be submitted to the well-known EGDI platform to ensure broader coverage and also a linkage to the JRC´s RMIS. The EGDI platform is also free of charge. See Fig 3. for the data storage overview.

The Zenodo repository services include data archiving with associated DOIs. Copyright licensing is also free of charge with Creative Commons. Costs related to gold open access for peer-reviewed publications are eligible as part of the Horizon Europe grant and are included in the project partner’s budget. The project web page is also one channel for project-related information. Data, however, is not stored on the project web page. Project´s social media channels are also sources of project-related information. The channels cover project-related activities, including links to publications and other open materials, but are not used for storing the project´s data.

The value and need for long-term public data preservation and related resources will be determined during the project’s progress.

The project coordinator Jari Joutsenvaara will lead the coordination of the updates to the data management plan. He will also organise data backup and storage, archiving and depositing the data to the Zenodo repository. He is also the link to the EGDI platform hosts for the final depositing of public data.

# 5 Data security

Intraproject data security is done by having reliable hosts (Microsoft Teams workspace provided; CSC, Finland) and supporting services. Access to services happens only through given user names and user login credentials. AGEMERA platform access is given only to specific users and to specific data sets, e.g., limited to certain trial areas or data sets generated with specific technologies.

Data security in the Zenodo repository is at a high level. The Zenodo data is stored at the servers in CERN Data Center with restricted access. All access to zenodo.org happens over HTTPS, and there is a network attack detection system. Zenodo stores user passwords using strong cryptographic password hashing algorithms. Both data and metadata files are kept in multiple online and independent replicas. Two independent checksums for each file enable automatic detection and recovery of data corruption on disks. In the improbable event that Zenodo will have to close operations, the data will be migrated to other relevant and suitable repositories. Moreover, as all uploads do have DOIs, all citations and links to Zenodo resources will not be affected.

Related to initially sensitive data, whether personal or IPR protected, the uploader shall ensure that such data is either anonymised to an appropriate degree or that full consent has been cleared. The uploading is done only by the data manager mentioned above. All data is reviewed, and quality is affirmed before submission to ensure an appropriate degree of anonymisation. IPR-protected data is not shared as open data.

# 6 Ethics

The AGEMERA project follows the EU and national ethics and data management standards. Article 14 of the Grant Agreement sets ethics and research integrity obligations that partners will follow. A separate Research ethics monitoring plan (D7.2.) is made covering the ethics issues raised by this project and measures that will be taken to ensure compliance with the ethical standards of the Horizon Europe programme.

# 7 Other issues

There are no other issues.

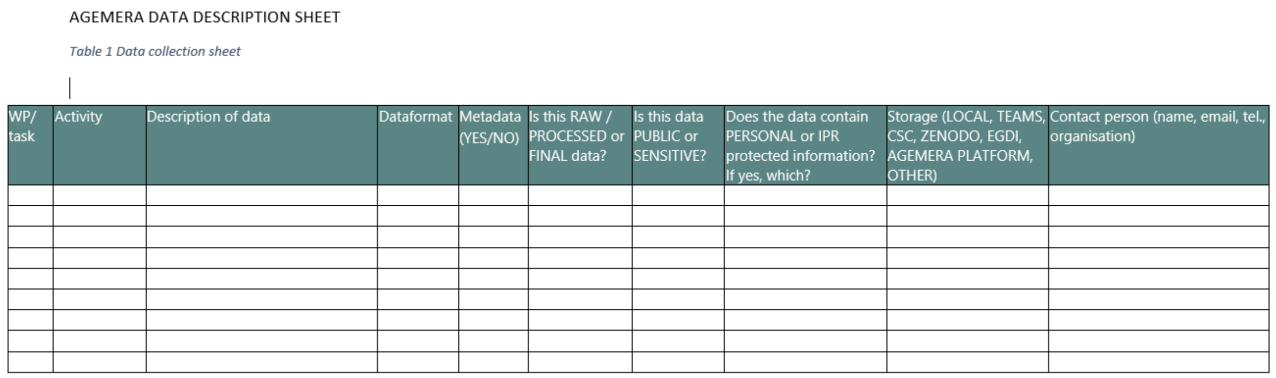
This project has received funding under the European Union's Horizon Europe research and innovation programme under grant agreement No 101058178.

# Attachments

## Query tables for the data management plan

In this document, the query tables are included as screenshot pictures. The actual tables are in the AGEMERA online workspace/ Channel WP7/ Folder Data Management Plan.

### Data description sheet



AGEMERA DATA DESCRIPTION SHEET

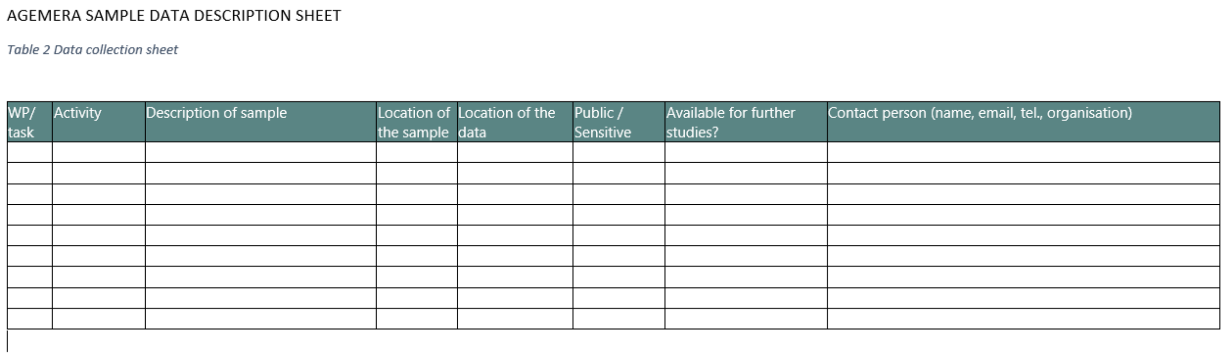
*Table 1 Data collection sheet*

| WP/ task | Activity | Description of data | Dataformat | Metadata (YES/NO) | Is this RAW / PROCESSED or FINAL data? | Is this data PUBLIC or SENSITIVE? | Does the data contain PERSONAL or IPR protected information? If yes, which? | Storage (LOCAL, TEAMS, CSC, ZENODO, EGDI, AGEMERA PLATFORM, OTHER) | Contact person (name, email, tel., organisation) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

The actual table is in the AGEMERA online workspace/ Channel WP7/ Folder Data Management Plan.

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

## Field study sample data sheet



AGEMERA SAMPLE DATA DESCRIPTION SHEET

*Table 2 Data collection sheet*

| WP/ task | Activity | Description of sample | Location of the sample | Location of the data | Public / Sensitive | Available for further studies? | Contact person (name, email, tel., organisation) |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

The actual table is in the AGEMERA online workspace/ Channel WP7/ Folder Data Management Plan.

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