

|  |  |
| --- | --- |
| D7.5 | Final Data Management Plan |
|  |
| **MAIL**: Identifying Marginal Lands in Europe and strengthening their contribution potentialities in a CO2 sequestration strategy |

MAIL project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 823805; [H2020 MSCA RISE 2018]

|  |  |
| --- | --- |
| Project title | Identifying Marginal Lands in Europe and strengthening their contribution potentialities in a CO2 sequestration strategy |
| Call identifier | H2020 MSCA RISE 2018 |
| Project acronym | MAIL |
| Starting date | 01.01.2019 |
| End date | 31.12.2021 |
| Funding scheme | Marie Skłodowska-Curie |
| Contract no. | 823805 |
|  |
| Deliverable no. | D7.5 |
| Document name | MAIL\_D7.5.pdf |
| Deliverable name | Data Management Plan |
| Work Package | WP7 |
| Nature[[1]](#footnote-1) | Report |
| Dissemination[[2]](#footnote-2) | CO |
| Editor | Charalampos Georgiadis, Petros Patias (AUTH) |
| Authors | Lampros Papalampros (HOMEOTECH) |
| Contributors | All consortium Partners |
| Date | 30.12.2021 |

# MAIL Consortium

|  |  |
| --- | --- |
| Aristotle University of Thessaloniki (AUTH) Greece | Industrieanlagen Betriebsgesellschaft MBH (IABG) Germany |
| Gounaris N. – Kontos K. OE (HOMEOTECH) Greece | Centrum Badan Kosmicznych Polskiej Akademii Nauk (CBK PAN) Poland  |
| Universitat Politecnica de Valencia (UPV) Spain | Fundacion Centro De Servicios Y Promocion FOrestral Y de su Industria De Castilla y Leon (CESEFOR) Spain |

# Abbreviations

|  |  |
| --- | --- |
| **Term** | **Explanation** |
| **AGB** | Above Ground Biomass |
| **AGBC** | Above Ground Biomass Carbon |
| **CA** | Consortium Agreement |
| **CSC Groups** | Carbon Sequestration Capacity Groups |
| **CO** | Coordinator |
| **DMP** | Data Management Plan  |
| **DoW** | Description of Work |
| **DR** | Deliverable Responsible |
| **EC** | European Commission |
| **EO** | Earth Observation |
| **EU** | European Union |
| **GA** | Grant Agreement |
| **GIS** | Geographical Information Systems |
| **IDM** | Impact & Dissemination Manager |
| **IPCC** | Intergovernmental Panel on Climate Change |
| **LULUCF** | Land Use, Land-Use Change & Forestry |
| **m/sm** | Mountainous/semi mountainous  |
| **MLs** | Marginal Lands |
| **MOOC** | Massive Online Open Course |
| **PM** | Project Manager |
| **QAP** | Quality Assurance Plan |
| **QM** | Quality Manager |
| **RS** | Remote Sensing |
| **SC** | Steering Committee |
| **SM** | Secondments Manager |
| **TC** | Technical Coordinator |
| **TL** | Task Leader |
| **WP** | Work Package |
| **WPL** | Work Package Leader |

Contents

[MAIL Consortium 3](#_Toc91756445)

[Abbreviations 4](#_Toc91756446)

[Executive Summary 10](#_Toc91756447)

[1. Introduction 10](#_Toc91756448)

[2. Datasets 11](#_Toc91756449)

[2.1 Hard Layer 11](#_Toc91756450)

[2.1.1 Data Summary 11](#_Toc91756451)

[2.1.1.1 Data set reference and name 12](#_Toc91756452)

[2.1.1.2 Data set description 12](#_Toc91756453)

[2.1.1.3 Name of person/organization responsible 12](#_Toc91756454)

[2.1.1.4 Nature and scale 12](#_Toc91756455)

[2.1.1.5 To whom might it be useful ('data utility')? 12](#_Toc91756456)

[2.1.2 Making data findable, including provisions for metadata 13](#_Toc91756457)

[2.1.2.1 Are the data produced and/or used in the project discoverable with metadata? 13](#_Toc91756458)

[2.1.2.2 Are data identifiable and locatable by means of a standard identification mechanism? 13](#_Toc91756459)

[2.1.3 Making data openly accessible 13](#_Toc91756460)

[2.1.3.1 Are datasets openly accessible? 13](#_Toc91756461)

[2.1.3.2 Is datasets access requiring some specific software tools? 13](#_Toc91756462)

[2.1.3.3 Where will the data and associated metadata, documentation and code be deposited 13](#_Toc91756463)

[2.1.3.4 Will data and all associated metadata be discoverable through catalogues and search engines? 13](#_Toc91756464)

[2.1.4 Making data interoperable 13](#_Toc91756465)

[2.1.4.1 Are the data produced in the project interoperable? 13](#_Toc91756466)

[2.1.4.2 Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability 13](#_Toc91756467)

[2.1.5 Increase data re-use 14](#_Toc91756468)

[2.1.5.1 Is the data safely stored in certified repositories for long term preservation and curation? 14](#_Toc91756469)

[2.1.5.2 How will the data be licensed to permit the widest re-use possible? 14](#_Toc91756470)

[2.1.5.3 When will the data be made available for re-use? 14](#_Toc91756471)

[2.1.5.4 Will data include provenance metadata to ensure full traceability of the product chain? 14](#_Toc91756472)

[2.1.5.5 Are data quality assurance processes described? 14](#_Toc91756473)

[2.1.5.6 Will be data full documented including all elements necessary to access, use, understand, and process, preferably via formal structured metadata? 14](#_Toc91756474)

[2.1.5.7 Will data be accessible via online services? 14](#_Toc91756475)

[2.1.5.8 Is the complete dataset available for download? 14](#_Toc91756476)

[2.1.6 Allocation of resources 15](#_Toc91756477)

[2.1.6.1 What are the costs for making data FAIR in your project? 15](#_Toc91756478)

[2.1.6.2 Who will be responsible for data management and preservation? 15](#_Toc91756479)

[2.1.6.3 Are the resources for long term preservation discussed? 15](#_Toc91756480)

[2.2 Final MLs 15](#_Toc91756481)

[2.2.1 Data Summary 15](#_Toc91756482)

[2.2.1.1 Data set reference and name 16](#_Toc91756483)

[2.2.1.2 Data set description 16](#_Toc91756484)

[2.2.1.3 Name of person/organization responsible 16](#_Toc91756485)

[2.2.1.4 Nature and scale 16](#_Toc91756486)

[2.2.1.5 To whom might it be useful ('data utility')? 16](#_Toc91756487)

[2.2.2 Making data findable, including provisions for metadata 17](#_Toc91756488)

[2.2.2.1 Are the data produced and/or used in the project discoverable with metadata? 17](#_Toc91756489)

[2.2.2.2 Are data identifiable and locatable by means of a standard identification mechanism? 17](#_Toc91756490)

[2.2.3 Making data openly accessible 17](#_Toc91756491)

[2.2.3.1 Are datasets openly accessible? 17](#_Toc91756492)

[2.2.3.2 Is datasets access requiring some specific software tools? 17](#_Toc91756493)

[2.2.3.3 Where will the data and associated metadata, documentation and code be deposited? 17](#_Toc91756494)

[2.2.3.4 Will data and all associated metadata be discoverable through catalogues and search engines? 17](#_Toc91756495)

[2.2.4 Making data interoperable 17](#_Toc91756496)

[2.2.4.1 Are the data produced in the project interoperable? 17](#_Toc91756497)

[2.2.4.2 Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability 17](#_Toc91756498)

[2.2.5 Increase data re-use 18](#_Toc91756499)

[2.2.5.1 Is the data safely stored in certified repositories for long term preservation and curation? 18](#_Toc91756500)

[2.2.5.2 How will the data be licensed to permit the widest re-use possible? 18](#_Toc91756501)

[2.2.5.3 When will the data be made available for re-use? 18](#_Toc91756502)

[2.2.5.4 Will data include provenance metadata to ensure full traceability of the product chain? 18](#_Toc91756503)

[2.2.5.5 Are data quality assurance processes described? 18](#_Toc91756504)

[2.2.5.6 Will be data full documented including all elements necessary to access, use, understand, and process, preferably via formal structured metadata? 18](#_Toc91756505)

[2.2.5.7 Will data be accessible via online services? 18](#_Toc91756506)

[2.2.5.8 Is the complete dataset available for download? 18](#_Toc91756507)

[2.2.6 Allocation of resources 19](#_Toc91756508)

[2.2.6.1 What are the costs for making data FAIR in your project? 19](#_Toc91756509)

[2.2.6.2 Who will be responsible for data management and preservation? 19](#_Toc91756510)

[2.2.6.3 Are the resources for long term preservation discussed? 19](#_Toc91756511)

[2.3 Marginal Lands classified in Carbon Sequestration Capacity Groups 19](#_Toc91756512)

[2.3.1 Data summary 19](#_Toc91756513)

[2.3.1.1 Data set reference and name 19](#_Toc91756514)

[2.3.1.2 Data set description 19](#_Toc91756515)

[2.3.1.3 Name of person/organization responsible 20](#_Toc91756516)

[2.3.1.4 Nature and scale 20](#_Toc91756517)

[2.3.1.5 To whom might it be useful ('data utility')? 20](#_Toc91756518)

[2.3.2 Making data findable, including provisions for metadata 20](#_Toc91756519)

[2.3.2.1 Are the data produced and/or used in the project discoverable with metadata? 20](#_Toc91756520)

[2.3.2.2 Are data identifiable and locatable by means of a standard identification mechanism? 21](#_Toc91756521)

[2.3.3 Making data openly accessible 21](#_Toc91756522)

[2.3.3.1 Are datasets openly accessible? 21](#_Toc91756523)

[2.3.3.2 Is datasets access requiring some specific software tools? 21](#_Toc91756524)

[2.3.3.3 Where will the data and associated metadata, documentation and code be deposited 21](#_Toc91756525)

[2.3.3.4 Will data and all associated metadata be discoverable through catalogues and search engines? 21](#_Toc91756526)

[2.3.4 Making data interoperable 21](#_Toc91756527)

[2.3.4.1 Are the data produced in the project interoperable? 21](#_Toc91756528)

[2.3.4.2 Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability 21](#_Toc91756529)

[2.3.5 Increase data re-use 21](#_Toc91756530)

[2.3.5.1 Is the data safely stored in certified repositories for long term preservation and curation? 21](#_Toc91756531)

[2.3.5.2 How will the data be licensed to permit the widest re-use possible? 22](#_Toc91756532)

[2.3.5.3 When will the data be made available for re-use? 22](#_Toc91756533)

[2.3.5.4 Will data include provenance metadata to ensure full traceability of the product chain? 22](#_Toc91756534)

[2.3.5.5 Are data quality assurance processes described? 22](#_Toc91756535)

[2.3.5.6 Will be data full documented including all elements necessary to access, use, understand, and process, preferably via formal structured metadata? 22](#_Toc91756536)

[2.3.5.7 Will data be accessible via online services? 22](#_Toc91756537)

[2.3.5.8 Is the complete dataset available for download? 22](#_Toc91756538)

[2.3.6 Allocation of resources 22](#_Toc91756539)

[2.3.6.1 What are the costs for making data FAIR in your project? 22](#_Toc91756540)

[2.3.6.2 Who will be responsible for data management and preservation? 22](#_Toc91756541)

[2.3.6.3 Are the resources for long term preservation discussed? 23](#_Toc91756542)

[3. Data Security 23](#_Toc91756543)

[4. Ethical Aspects 23](#_Toc91756544)

[4.1 Misuse 23](#_Toc91756545)

[4.2 Data 24](#_Toc91756546)

[4.3 Other issues 24](#_Toc91756547)

[4.3.1 Compliance with H2020 Open Research Data Pilot 24](#_Toc91756548)

[4.3.1.1 Data generated in the project are available in a repository with proper metadata: 24](#_Toc91756549)

[4.3.1.2 Data generated in the project are available free-of-charge: 25](#_Toc91756550)

[4.3.1.3 Scientific results (e.g. publications) of the project are (data are available free-of-charge, information on required tools and instruments are provided): 25](#_Toc91756551)

# Executive Summary

Deliverable 7.5 Final Data Management Plan (DMP) provides the general strategy for MAIL data and information management. The final DMP provides a general overview of the type of data that was generated during the project and explains how the data (including metadata) were collected, stored and made accessible.

The DMP was planned to be a living document that was updated regularly throughout the project’s lifetime. **The present document is the final version of the DMP**, where the final datasets that produced within the project activities are described, as well as the storing, accessing and documentation policies to be applied to them.

# Introduction

Marginal Lands (MLs) could offer suitable carbon sinks by afforestation/reforestation projects without being competitive to food production, conforming to EU/Global policies. The main objective of ***MAIL*** is to trigger utilization of MLs as Carbon Sinks by activities related to forestry and agriculture. The project aims to detect and classify mountainous and semi-mountainous Marginal Lands (m/sm MLs), in order to deliver a web-based regional platform with data, methodology and applications which will be valuable for policy makers, stakeholders or researchers.

***MAIL*** assess MLs’ presence, distribution and suitability in smart and efficient way, considering that marginality differs within European bioclimatic zones. It detects MLs through multiple layer analysis of existing EU/Global scale datasets (land cover, land use, soil or climatic information, etc.) and examines, validates and classifies MLs, through field stratified sampling, into Carbon sequestration capacity categories.

In pilot areas ***MAIL*** augments the achieved accuracy by using enhanced spatial and radiometric resolution of full free and open access satellite datasets. Furthermore, it proposes actions that have to be taken in order to increase Carbon sequestration capacity and it will evaluate their feasibility by using a set of criteria for accurate cost estimation. Finally, it delivers the developed methodologies, algorithms and techniques on a web-application platform for on-demand production of thematic maps within Europe.

***MAIL*** consists of a multidisciplinary Consortium of 6 experienced partners, from 4 EU Member States, and gathers the expertise required for identifying, mapping and assessing MLs as a potential Carbon Sinks.

All the R&I activities were implemented through research personnel secondments, fully founded by Horizon 2020/ MSCA-RISE 2018.

The purpose of the Final Data Management Plan (DMP) is to provide an analysis of the main elements of the data management that will be generated by the project.

This deliverable sets out a general strategy for ***MAIL*** information management. The Final DMP provides an overview of the type of data generated during the project and explains how the data (including metadata) is collected, stored and made accessible.

The present document is the final version of the DMP. It is generated at the end of the project activities (M36); therefore, it outlines the datasets that were produced within the project and the storing, accessing and documentation policies to be applied to them. The Final DMP was intended to be a living document that was updated as the project progresses, and it improved the accomplishment of FAIR principles.

# Datasets

Within the MAIL project, the proposed methodology is divided into two main steps: The implementation of the “hard” thresholds and constrains methodology which relies on the exclusion of areas that do not meet the requirements of the definition of Marginal Lands (i.e. agricultural lands, forest and impervious areas, permanent water and snow areas, peatbogs, marshes, and more) and the “additional indicators/thresholds” phase based on the development of specific sets of additional indicators (i.e. “soft” constrains) that will be applied according to the physical characteristics of each selected test site. These indicators are considering additional variables including the terrain and soil characteristics of an area and soil productivity constraints.

## Hard Layer

### Data Summary

In the first phase, a top-down stepwise approach is followed, in which areas that are not MLs are incrementally removed, based on thresholds of various marginality criteria and indicators. These results are an intermediate layer, named “ML HARD” including all potential MLs after the exclusion of all LULC types that are not fulfilling the definition of MLs ***(DOI: 10.5281/zenodo.5808943).***

#### Data set reference and name

***HARD-xxx.tif*** or files based on OGC standards*.*

#### Data set description

As described in D2.3, the first phase of the proposed methodology (“hard” thresholds phase) suggests the delineation of specific Land Use/Land Cover (LULC) classes that should be excluded from further analysis and that are not MLs according to the literature review and the definition of Marginal Lands. Namely, five different types of LULC should be identified and incrementally removed: Other Land Cover areas including Water bodies, Permanent Snow, Marshes and Peatbogs, Forest Areas, Croplands, Impervious Areas, Protected Areas, and Changed Areas. To identify these categories and to delineate those LULC, several combinations of the available datasets as already defined in Task 2.1 were tested. Considering the spatial resolution, and the reference year of each available dataset, various analysis tools (e.g., Fuzzy Overlay, Weighted Overlay, etc.) were used to combine them. To perform those functions, an initial validation of each available dataset was also implemented by combining the existing datasets with historical high -resolution satellite imagery (e.g., google earth images).

All layers’ original resolution were resampled to the finest dataset resolution (i.e. 10m) of the S2GLC (Land Cover Map of Europe) which is used as Basemap. Moreover, to combine the datasets used in this first phase of “hard” thresholds, all the produced intermediate raster layers, are converted into Binary Raster (i.e., 0 = no data, 1 = data that need to be excluded and they don’t represent “marginal lands”). Furthermore, to avoid errors due to mismatches during the implementation of the proposed algorithms, the initial Basemap is set as “Snap Raster” Layer.

#### Name of person/organization responsible

AUTH (Greece).

#### Nature and scale

Pixel size 10\*10m for all marginal lands of Europe, as detected in D2.3 of ***MAIL.***

#### To whom might it be useful ('data utility')?

Students, Land owners, stakeholders, researchers, agencies, institutes, e.g. EFI, EIONET of EEA, IUFRO etc.

### Making data findable, including provisions for metadata

#### Are the data produced and/or used in the project discoverable with metadata?

Files with legend (thematic maps, etc.) will be provided at Zenodo open-access repository. Efforts will be done towards including metadata compatible with INSPIRE directive.

#### Are data identifiable and locatable by means of a standard identification mechanism?

Zenodo connects each submission with a persistent digital object identifier (DOI), which makes the stored items easily citeable.

### Making data openly accessible

#### Are datasets openly accessible?

Yes, they are openly accessible.

#### Is datasets access requiring some specific software tools?

No, it could be opened by any GIS software (QGIS, ArcGIS).

#### Where will the data and associated metadata, documentation and code be deposited

At Zenodo open-access repository.

#### Will data and all associated metadata be discoverable through catalogues and search engines?

The data is interoperable.

### Making data interoperable

#### Are the data produced in the project interoperable?

The data is interoperable.

#### Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability

The data use standard vocabularies.

### Increase data re-use

#### Is the data safely stored in certified repositories for long term preservation and curation?

A local repository is deployed for project needs. Finalized version will be stored in certified repositories. As already mentioned for that reason Zenodo open-access repository will be used.

#### How will the data be licensed to permit the widest re-use possible?

Dataset usage will require citation of the dataset source.

#### When will the data be made available for re-use?

After its upload to Zenodo, close to project’s end.

#### Will data include provenance metadata to ensure full traceability of the product chain?

Yes.

#### Are data quality assurance processes described?

Yes. Internally the repository has been set to a cloud server as “*read only*” to avoid unwanted editing. At Zenodo repository data quality assurance processes held by Zenodo’s procedures.

#### Will be data full documented including all elements necessary to access, use, understand, and process, preferably via formal structured metadata?

Yes, metadata will be structured and understandable.

#### Will data be accessible via online services?

Yes, through the Zenodo platform.

#### Is the complete dataset available for download?

It will be available at the end of the project at Zenodo repository, when all the related tasks will be completed Furthermore, the open access datasets used for the implementation of the project can be downloaded from their original sources as described in D2.2.

### Allocation of resources

#### What are the costs for making data FAIR in your project?

Not defined.

#### Who will be responsible for data management and preservation?

As already stated, the final data will be uploaded to Zenodo, following Zenodo’s backup plan. AUTH will held an original copy at its automated backup processes, as described at §3. Additionally, if any future changes will come up, then they will be uploaded to Zenodo as version 2 by the Project Coordinator.

#### Are the resources for long term preservation discussed?

Data will be downloadable per request at the project website that will be active for 5 years after project ends.

## Final MLs

### Data Summary

As described in D2.3, these datasets represent all the detected MLs in EU 28, with a weighted value per pixel. This value based on a Pairwise Comparison Matrix (PCM) of ranks where the value of each indicator (i.e., slope, moisture, salinity, productivity, soil contamination, pH etc.) are grouped into 3 classes according to their physical characteristics. This result in a weight value per pixel in EU 28. The final step in the mapping of marginal lands, is the reclassification of the resulting product of the weighted overlay into 3 classes, depicting marginality: *1)* *Marginal lands with high plantation suitability, 2) Marginal lands with low plantation suitability* and *3) Potentially unsuitable lands*. The subdivision of these categories was carried out with three different methodological approaches;

The first is to calculate the maximum and minimum value obtained in the MLs layer and divide the difference by three. This will obtain three ranges of equal magnitude.

The second methodological approach is to calculate the 25th and 75th percentiles of the values obtained in the MLs layer. This approach would penalize the “Marginal lands with high plantation suitability” and the "Potentially unsuitable lands" and increasing the number of pixels belonging to the middle layer "Marginal lands with low plantation suitability".

The third one is to calculate the 33rd and 66th percentile of the values obtained in the layer MLs. In this case, the total number of pixels is divided equally into the three categories. But with the difference with respect to the a) method that in this case the thresholds are established with the maximum number of pixels and not with the maximum and minimum values ***(DOI: 10.5281/zenodo.5809011).***

#### Data set reference and name

***MLs\_Europe \_Div3\_xx.tif*** files based on OGC standards. This is one of the final datasets that will depict from the geo-analysis of all the indicators and represents the three ranges in equal magnitude.

***MLs\_Europe \_p2575\_xx.tif*** files based on OGC standards. This is one of the final datasets that will depict from the geo-analysis of all the indicators and represents the three ranges in 25th and 75th percentiles of the values obtained.

***MLs\_Europe \_p3366\_xx.tif*** files based on OGC standards. This is one of the final datasets that will depict from the geo-analysis of all the indicators and represents the three ranges in 33th and 66th percentiles of the values obtained.

#### Data set description

As already described, this is the second phase of the analysis that has to do with the ranking of the already selected MLs of the first phase. The resulting potential MLs are further downsized using additional indicators, related to Terrain, Soil, Sustainability, Climate and Productivity thresholds. Finally, MLs are classified into 3 classes depicting marginality: 1) Marginal lands with high plantation suitability, 2) Marginal lands with low plantation suitability and 3) Potentially unsuitable lands.

#### Name of person/organization responsible

AUTH (Greece).

#### Nature and scale

Pixel size 10\*10m for all marginal lands of Europe, as detected in D2.3 of ***MAIL.***

#### To whom might it be useful ('data utility')?

Students, Land owners, stakeholders, researchers, agencies, institutes, e.g. EFI, EIONET of EEA, IUFRO etc.

### Making data findable, including provisions for metadata

#### Are the data produced and/or used in the project discoverable with metadata?

Files with legend will be provided at Zenodo open-access repository. Efforts will be done towards including metadata compatible with INSPIRE directive.

#### Are data identifiable and locatable by means of a standard identification mechanism?

Zenodo connects each submission with a persistent digital object identifier (DOI), which makes the stored items easily citeable.

### Making data openly accessible

#### Are datasets openly accessible?

Yes, they are openly accessible.

#### Is datasets access requiring some specific software tools?

No, it could be opened by any GIS software (QGIS, ArcGIS).

#### Where will the data and associated metadata, documentation and code be deposited?

At Zenodo open-access repository.

#### Will data and all associated metadata be discoverable through catalogues and search engines?

The data is interoperable.

### Making data interoperable

#### Are the data produced in the project interoperable?

The data is interoperable.

#### Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability

The data use standard vocabularies.

### Increase data re-use

#### Is the data safely stored in certified repositories for long term preservation and curation?

A local repository is deployed for project needs. Finalized version will be stored in certified repositories. As already mentioned for that reason Zenodo open-access repository will be used.

#### How will the data be licensed to permit the widest re-use possible?

Dataset usage will require citation of the dataset source.

#### When will the data be made available for re-use?

After its finalisation, close to project’s end.

#### Will data include provenance metadata to ensure full traceability of the product chain?

Yes.

#### Are data quality assurance processes described?

Yes. Internally the repository has been set to a cloud server as “read only” to avoid unwanted editing. At Zenodo repository data quality assurance processes held by Zenodo’s procedures.

#### Will be data full documented including all elements necessary to access, use, understand, and process, preferably via formal structured metadata?

Yes, metadata will be structured and understandable.

#### Will data be accessible via online services?

Yes, through the Zenodo platform.

#### Is the complete dataset available for download?

It will be available at the end of the project at Zenodo repository, when all the related tasks will be completed. Furthermore, the open access datasets used for the implementation of the project can be downloaded from their original sources as described in D2.2.

### Allocation of resources

#### What are the costs for making data FAIR in your project?

Not defined.

#### Who will be responsible for data management and preservation?

As already stated, the final data will be uploaded to Zenodo, following Zenodo’s backup plan. AUTH will held an original copy at its automated backup processes, as described at §3. Additionally, if any future changes will come up, then they will be uploaded to Zenodo as version 2 by the Project Coordinator.

#### Are the resources for long term preservation discussed?

Data will be downloadable per request at the project website that will be active for 5 years after project ends.

## Marginal Lands classified in Carbon Sequestration Capacity Groups

### Data summary

Previously detected m/sm MLs was further classified in Carbon Sequestration Capacity (CSC) Groups categories. The methodology based on multicriteria GIS analysis with data including tree species maps, land cover maps and Aboveground Biomass maps. The aim was to estimate potential suitable species for afforestation for each Marginal Land as well species’ Above Ground Biomass Carbon (AGBC) and proceed to classification into CSC groups ***(DOI: 10.5281/zenodo.5808901).***

#### Data set reference and name

***CSC\_Groups.tif*** based on OGC standards*.*

#### Data set description

The CSC of MLs was estimated based on potential suitable species for afforestation and their Aboveground Biomass Carbon. The MLs, as calculated on Task 2.3, are the basemap, where the most frequent species from neighbor forested areas, both dominant 1 and 2 species, and species’ Aboveground Biomass Carbon values are assigned. Dominant 1 and 2 species of neighbor forested areas are adapted to the ecological and climatological conditions and therefore are considered to be the most suitable for afforestation projects. Through classification into CSC groups, we get a better understanding regarding the relative interconnections between groups and each one's potential carbon stock.

The frequency distribution of the formula’s results is presented in a histogram. Classification into CSC groups was done by manually defining classes ranges, in such a way so each class to cover approximately the same area across Europe, with the exception of higher and lower sequestration groups, Group A and Group E respectively. Group A represents higher sequestration MLs, covering 5% of Europe’s total MLs and on the other side Group E represents lower sequestration MLs covering 31% of Europe’s MLs. The final classification is presented in the following relative pie chart.

Figure 1: Groups’ distribution in percentages for European cover

#### Name of person/organization responsible

Cesefor (Spain).

#### Nature and scale

Pixel size 10\*10m for all marginal lands of Europe, as detected in D2.3 of ***MAIL.***

#### To whom might it be useful ('data utility')?

Students, Land owners, stakeholders, policy makers, researchers, agencies, institutes, e.g. EFI, EIONET of EEA, IUFRO etc.

### Making data findable, including provisions for metadata

#### Are the data produced and/or used in the project discoverable with metadata?

Files with legend will be provided at Zenodo open-access repository. Efforts will be done towards including metadata compatible with INSPIRE directive.

#### Are data identifiable and locatable by means of a standard identification mechanism?

Zenodo connects each submission with a persistent digital object identifier (DOI), which makes the stored items easily citeable.

### Making data openly accessible

#### Are datasets openly accessible?

Yes, they are openly accessible.

#### Is datasets access requiring some specific software tools?

No, it could be opened by any GIS software (QGIS, ArcGIS).

#### Where will the data and associated metadata, documentation and code be deposited

 At Zenodo open-access repository.

#### Will data and all associated metadata be discoverable through catalogues and search engines?

The data is interoperable.

### Making data interoperable

#### Are the data produced in the project interoperable?

The data is interoperable.

#### Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability

The data use standard vocabularies.

### Increase data re-use

#### Is the data safely stored in certified repositories for long term preservation and curation?

A local repository is deployed for project needs. Finalized version will be stored in certified repositories. As already mentioned for that reason Zenodo open-access repository will be used.

#### How will the data be licensed to permit the widest re-use possible?

Dataset usage will require citation of the dataset source.

#### When will the data be made available for re-use?

After its upload to Zenodo, close to project’s end.

#### Will data include provenance metadata to ensure full traceability of the product chain?

Yes.

#### Are data quality assurance processes described?

Yes. Internally the repository has been set to a cloud server as “read only” to avoid unwanted editing. At Zenodo repository data quality assurance processes held by Zenodo’s procedures.

#### Will be data full documented including all elements necessary to access, use, understand, and process, preferably via formal structured metadata?

Yes, metadata will be structured and understandable.

#### Will data be accessible via online services?

Yes, through the Zenodo platform

#### Is the complete dataset available for download?

It will be available at the end of the project at Zenodo repository, when all the related tasks will be completed.

### Allocation of resources

#### What are the costs for making data FAIR in your project?

Not defined.

#### Who will be responsible for data management and preservation?

As already stated, the final data will be uploaded to Zenodo, following Zenodo’s backup plan. AUTH will held an original copy at its automated backup processes, as described at §3. Additionally, if any future changes will come up, then they will be uploaded to Zenodo as version 2 by the Project Coordinator.

#### Are the resources for long term preservation discussed?

Data will be downloadable per request at the project website that will be active for 5 years after project ends.

# Data Security

The local repository backup will be scheduled on a daily basis using an external hard drive (drive 1). The backup plan is based on full and incremental backup methods.

A full backup will be performed once a month and will store all the repository data. This backup will be the base for the daily incremental backups. That incremental backups will store changes to the data against the latest backup.

Every other month the external hard drive will change to another one (drive 2). The process will be repeated for another month, while the drive 1 will be stored in a safe place. At the third month drive 1 will be emptied and take the place of drive 2 for another set.

Every six months and for the duration of one year a full back up will be kept in another external drive (drive 3).

That way drive 1 & 2 will always contain full and incremental backups of the last two months, while drive 3 will enclose two full backups of the last two semesters.

# Ethical Aspects

## Misuse

The main research findings of the project will be raster or vector files of EU 28 classifying the marginal lands based on their Carbon sequestration capacity groups. The produced data (classification maps) will only provide information about the type and capacity of the marginal lands. As such they don’t have the potential to harm humans, animals, or the environment. In addition, such information cannot be channelled into crime or terrorism. The research findings will be derived using free and open research data and will be offered through the project’s website. The research findings will be offered as is.

With respect to personal data. The project platform will offer registration for 4 different services namely:

1. Newsletter registration
2. Virtual room registration
3. MOOC registration
4. Open research data pilot registration

During the user registration the platform will comply with the General Data Protection Regulation (GDPR). The collected data will be used only for the purposes of the project and only if the users consent to that. The project platform will be hosted by the University’s IT services which is certified and has all the required security features to comply with GDPR. As a result, the personal data protection will fall under the umbrella of AUTH’s IT services.

Thus, there will be no misuse of the research findings by the partners.

## Data

The Coordinator (AUTH) confirms that a Data Protection Officer (DPO) has been appointed by the organization. The role of DPO is to monitor the GDPR compliance of the whole AUTH data.

The consortium processed freely available data like satellite images (Sentinel, LANDSAT, etc), land cover datasets like CORINE, Global Climate data, Natura 2000, etc. along with national/ regional datasets collected for the four pilot area sites. The data was processed in order to define Marginal lands and estimate their potential CO2 sequestration based on different criteria. WP2 “Research” has dedicated tasks that studied and investigated the current state of the art and proposed the most relevant data that was used for the successful implementation of the project.

No personal data is involved.

## Other issues

### Compliance with H2020 Open Research Data Pilot

#### Data generated in the project are available in a repository with proper metadata:

✓Yes

Partially: \_% of datasets are available in a repository with proper metadata

No

Comment/Justification: Data will be on repositories from the partners. The consortium will consider to store them on a common repository.

#### Data generated in the project are available free-of-charge:

✓Yes

Partially: \_\_% of datasets are available free-of-charge

No

Comment/Justification: In some cases, downloading the data will be under request, but free-of-charge. Visualization and consultation will be always free.

#### Scientific results (e.g. publications) of the project are (data are available free-of-charge, information on required tools and instruments are provided):

✓ Yes, they are

Partially: \_\_% of scientific results are reproducible

No

Comment/Justification:

The results that are coming from workflows are reproducible

1. **R** = Report, **P** = Prototype, **D** = Demonstrator, **O** = Other [↑](#footnote-ref-1)
2. **PU** = Public, **PP** = Restricted to other programme participants (including the Commission Services), **RE** = Restricted to a group specified by the consortium (including the Commission Services), **CO** = Confidential, only for members of the consortium (including the Commission Services). [↑](#footnote-ref-2)