

# GOLD

Growing energy crops on contaminated  
land for biofuels and soil remediation

 Ref. Ares(2023)5393936 - 03/08/2023



## Data Management Plan



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 101006873.

## Document Summary

Deliverable Number: **D5.8**

Version: 1.0

Due date: 31.10.2021

Actual submission date: 17.02.2022

Work Package: 5 - Coordination & management

Task: 5.5: Data Management and exploitation of the results

Lead beneficiary: ETA

Editors/Authors: Emma Fromant, Maurizio Cocchi

Dissemination level: Public

### a. Document history

| Version | Date       | Beneficiary | Author/Reviewer |
|---------|------------|-------------|-----------------|
| 1.0     | 17.02.2022 | ETA         | Emma Fromant    |

Project Number: **101006873**

Grant Agreement Number: **101006873**

Programme: **GOLD**

Start date of Project: **1 May 2021**

Duration: **48 months**

Project coordinator: **CRES**

This project has used a standard methodology already developed in BIKE project (Grant Agreement number: 952872), following EU recommendations. Ad hoc modifications were added to comply with the Grant Agreement conditions for GOLD (Grant Agreement number: 101006873).

## Partners

|   |
|---|
| CRES - Centre for Renewable Energy Sources and Saving Foundation, Greece                                |
| AUA – Geoponiko Panepistimion Athinon, Greece   |
| TUM - Technische Universität München, Germany   |
| RE-CORD - Consorzio per la Ricerca e la Dimostrazione sulle Energie Rinnovabili, Italy                  |
| ETA - ETA Energia, Trasporti, Agricoltura, Italy  |
| Uni-Lublin - Uniwersytet Marii Curie-Sklodowskiej, Poland   |
| TNO - Nederlandse Organisatie Voor Toegepast Natuurwetenschappelijkonderzoek<br>TNO, Netherlands        |
| CERTH - ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS, Greece                                      |
| UNIBO - Alma Mater Studiorum - Università di Bologna, Italy   |
| INRAE - Institut National de Recherche pour l'Agriculture, l'Alimentation et<br>l'Environnement, France |
| YNCREA HDF – Junia, France  |
| UNL - Universidade Nova de Lisboa, Portugal   |
| ICL - Imperial College of Science Technology and Medicine, United Kingdom                               |
| WR - Stichting Wageningen Research, Netherlands   |
| METE S.A. - METE AE METALLEFTIKI EMPORIKI TEHNIKI AE*MINING TRADING<br>TECHNICAL SA, Greece             |
| HUNAN - Hunan Agricultural University, China  |
| UDES - Université de Sherbrooke, Canada   |
| IBFC - Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, China                   |

### Statement of Originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

### Disclaimer of warranties

The sole responsibility for the content of this report lies with the authors. It does not necessarily reflect the opinion of the European Union. Neither the European Commission nor INEA are responsible for any use that may be made of the information contained therein.

## Table of contents

|   |           |
|---|-----------|
| <b>Document Summary</b> .....                               | <b>2</b>  |
| a. Document history .....                                   | 2         |
| <b>Partners</b> .....                                       | <b>3</b>  |
| <b>Table of contents</b> .....                              | <b>4</b>  |
| <b>Introduction</b> .....                                   | <b>5</b>  |
| <b>1 Scope and objectives</b> .....                         | <b>5</b>  |
| <b>2 GOLD data summary</b> .....                            | <b>5</b>  |
| <b>3 FAIR data</b> .....                                    | <b>7</b>  |
| <b>4 Making data findable</b> .....                         | <b>7</b>  |
| 4.1 The Horizon Community in ZENODO .....                   | 7         |
| 4.2 Metadata .....  | 7         |
| 4.3 Approach to search keywords.....                        | 8         |
| 4.4 Naming conventions .....                                | 8         |
| <b>5 Making data accessible</b> .....                       | <b>8</b>  |
| <b>6 Making data interoperable</b> .....                    | <b>9</b>  |
| <b>7 Increasing data re-use (licensing)</b> .....           | <b>9</b>  |
| <b>8 Allocation of resources</b> .....                      | <b>9</b>  |
| <b>9 Rights to access and re-use of research data</b> ..... | <b>10</b> |
| <b>10 Data sets overview</b> .....                          | <b>10</b> |
| <b>Annex I – Data Collection Form</b> .....                 | <b>12</b> |
| <b>Annex II – Data Inventory Table</b> .....                | <b>13</b> |

## Introduction

This document outlines the Data Management Plan (DMP) for GOLD project, which will be used as a support throughout the project and updated in M48 at the close of the project to support the project findings in the future.

## 1 Scope and objectives

The scope of this documents is to outline the procedures to handle project data, in regards to its collection, organisation, and licensing, as well how GOLD will meet the requirements of the open access mandate. In addition, it provides an indication of how the data will be preserved after the project is completed and it's benefit on society.

The GOLD DMP objective is to provide the following information:

- The Data Management Policy agreed upon by the consortium.
- the updated list of data (datasets) to be generated or collected so far and in the future, as well as information on beneficiaries of such data, during the project and after its closure.
- standards and metadata at the basis of each dataset, as the enabling data to be found and understood correctly, (ideally) according to the specific standards currently in use in the scientific disciplines.
- the policy on Data Sharing: as sharing data outside the project team is the default, legitimate reasons for not sharing resulting data will be duly explained in the DMP (e.g., third-party copyright issues, confidentiality, IPR rights or when open dissemination could put at risk project results);
- the procedures for Data archiving and preservation: since usability of data depends not only on storage and backup but also on well-preserved software and on conversion to new file formats, it will be described where data, metadata, documentation, and software will be preserved for the long-term, shared and made open for re-use.

It is intended that the above data will be generated by the partners and shared and gathered throughout the project. Such information will be collected periodically throughout the project by filling the form detailed in Annex I and will be summarised and included in the final version of the plan released in M48 of the project.

## 2 GOLD data summary

GOLD is an international collaboration between Europe, China and Canada, spanning 12 different countries in total. As such the project is expected to generate and collect a vast amount of research data in the form of numerical, textual, modelling and mapping data, focusing on three pillars:

1. The optimisation of selected energy crops for phytoremediation for soil decontamination purposes.
  2. Conversion of contaminated feedstocks into clean biofuels based on two conversion routes; gasification and pyrolysis.
  3. Assessment of the value chain for biofuels production and land remediation through modelling and analysis of costs, sustainability, and SDGs.
-

Due to continuously evolving technologies, it is important that data is generated using the following principles are considered when selecting the file format; that it is commonly used, it has open specifications, it is independent of specific software, developers, or suppliers.

Table 1<sup>1</sup> below lists the type of data anticipated to be generated, as well as the suggested formats to be used for processing, sharing and preservation.

*Table 1 Summary of data formats*

| Type of data   | Formats used during data processing  | Formats for sharing reuse and preservation   |
|--|--|--|
| Numerical or textual tabular data  | Microsoft Excel (.xls/.xlsx)   | Comma-separated values (.csv)  |
| Qualitative textual data   | Microsoft Word (.doc/.docx)  | Rich Tech File (.rtf)<br>Non-Unicode TXT (.txt)<br>OpenDocument Text (.odt)<br>PDF/A (.pdf)<br>Txt file (.shp format)  |
| Numerical or qualitative textual data  | Acrobat Adobe Reader (.pdf)  | PDF/A (.pdf)   |
| Spreadsheets   | OpenDocument Spreadsheet (.ods)<br>Comma Separated Values (.csv)                           | MS Excel (.xls, .xlsx)<br>PDF/A (.pdf)<br>OOXML (.docx, .docm)   |
| Topic modelling data   | Mallet format (.mallet)  | Comma-separated values (.csv)  |
| Simulation model data  | Text model format (.mdl)   | The mathematical model could be saved using standard differential equations symbols in .rtf, .csv and .txt files in case the simulation model is developed using a proprietary software. Simulated values will be saved as numerical data.                         |
| Statistical tabular data   | STATA format (.dta)  | Comma-separated values (.csv)<br>STATA format (.dta)<br>MS Excel (.xls)  |
| Geospatial vector data format for geographic information system (GIS) software | Shape GIS format (.shp – shape format, .shx – shape index format, .dbf – attribute format) | Shapefiles can support point, line, and area features. Area features are represented as closed loop, double-digitized polygons. Attributes are held in a dBASE® format file. Each attribute record has a one-to-one relationship with the associated shape record. |
| Audio data   | mp3 format (.mp3)  | Audio recordings will be deleted after their transcription and only the processed transcripts will be shared and preserved.  |
| Images (raster)  | JPEG (.jpg, .jpeg)<br>TIFF (.tif, .tiff)<br>PNG (.png)                                     | JPEG 2000 (.jp2)   |
| Images (vector)  | scalable vector graphics (.svg)  | Adobe Illustrator (.ai)<br>PostScript (.eps)   |

<sup>1</sup> As it is not always possible to select formats that meet with all the ideal attributes, a number of file formats has been listed as suggestions. This list could change over time as new formats will be developed and others will fall into disuse.

Annex II “Data inventory table” lists in detail the deliverables and other relevant data produced during the activities of GOLD. The inventory table summarizes the key information regarding each resource such as type, format, methodology, date of availability, etc. This table will be a summary of the information gathered from partners via the filling of Annex I.

## **3** FAIR data

This DMP, coherently with the EU guidelines, describes the data management procedures according to FAIR (findable, accessible, interoperable, and re-usable) data management.

Commonly identified as the four main features of project research data, they are relevant for allowing their maximum knowledge circulation as well as final return on investment (“scientific ROI”). This is in line with the indications provided by the European Council<sup>2</sup>.

## **4** Making data findable

GOLD research project is expected to generate and collect vast amounts of data. In order to make this data manageable, data with a similar focus and scope will be grouped together to form datasets.

### **4.1 The Horizon Community in ZENODO**

This DMP recommends that GOLD uses the ZENODO repository as the main tool to make data findable in accordance with the Open Access Mandate.

Through ZENODO a GOLD community can be established on the website, and all public datasets, deliverables and scientific publications can be uploaded. In addition, uploads to ZENODO can be linked easily to OpenAIRE, which will maximise findability. All uploads will be enriched with standard ZENODO metadata, including Grant Number, Project Acronym, version control and automatically assigned DOIs to all uploaded elements.

### **4.2 Metadata**

Each dataset will be described using standard metadata – such a Dublin Core and DataCite Metadata Schema, coherently to what entailed by the OpenAIRE guidelines, as well as what indicated by the Research Data Alliance (RDA) on the Metadata Standards Directory<sup>3</sup>.

In addition, metadata associated with each published data set in ZENODO will by default be as follows:

- Digital Object Identifiers
- Version numbers
- Bibliographic information

---

<sup>2</sup> [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

<sup>3</sup> <https://www.rd-alliance.org/groups/metadata-standards-directory-working-group.html>

---

- Keywords
- Abstract/description
- Associated project and community
- Associated publications and reports
- Grant information
- Access and licensing info
- Language

### 4.3 Approach to search keywords

All partners will be responsible for uploading public datasets that they have generated and to assign specific keywords relevant to these datasets. Dataset specific keywords must be descriptive to the content of the dataset. In addition, this DMP recommends a set of general keywords that should apply to all public datasets, scientific publications, and public deliverables for GOLD project. These are as follow:

Biofuels, contaminated land, phytoremediation, low ILUC, energy crops

### 4.4 Naming conventions

Datasets should also be named in a consistent manner to permit data visibility, discoverability, citation and permanent online tracking. This data management plan recommends that document names should consist of the following information:

Date: Project Acronym: WP title or description specifying WP aims: Task or Deliverable number: Task title or description: Version number: Beneficiary.

Example:

DDMMYYYY\_GOLD\_TaskTitle\_DX.X\_v.v\_BENEFICIARY

## 5 Making data accessible

The GOLD consortium will follow the H2020 Open Access Mandate of access to research data, ensuring that the results will contribute to future works of the European research community. This will apply to deliverables, and final research data. All public datasets, scientific publications and deliverable will be uploaded to a repository to be made openly available and free of charge. The majority of GOLDs deliverables are public but there may be a possibility that some datasets might be restricted due to protection for commercial exploitation. If such cases arise during the project, this will be informed in the final version of the DMP.

The data repository for the project chosen by GOLD partners will be an institutional or public repository:

- fully compliant with long-term preservation and attribute persistent unique identifiers to the archived datasets (such as Handle or DOI);
- supportive of open licenses and different access levels;

capable to adopt descriptive metadata standards as required by the OpenAIRE Guidelines and allow cross-linking between publications and relevant datasets, guaranteeing their compatibility with OpenAIRE requirements.

---



It is recommended that Zenodo is the chosen repository for GOLD. Zenodo's functionalities are: a full interface enabling linking research outputs to datasets and funding information; all open content is harvestable via OAI-PMH by third parties; all the different versions of a file can be supported by a top-level DOI; data is stored at CERN, counting on considerable experience operating large scale digital repositories; data files and metadata are kept in multiple online and offline copies; research materials can set to share with reviewers only, and also embargoed.

Partners might still choose to deposit their dataset/s in more than one repository (e.g. in a multi-disciplinary, wide known repository such as Zenodo to comply with OpenAIRE requirements and in their institutional repository for their usual visibility and dissemination among other researchers), however the data should be linked to Zenodo.

## 6 Making data interoperable

All datasets will be described using standard descriptive metadata, such as Dublin Core and DataCite Metadata Schema, for ensuring metadata interoperability for indexing and discoverability. All relevant documentation explaining codebooks, users' manuals, data collection procedures and analysis will be made available along with the data, in compliance with requirements of intelligibility, reproducibility and validation of project findings. Variable names of data derived from other official sources (e.g. Eurostat, others..), will be consistent with the original source names.

## 7 Increasing data re-use (licensing)

In GOLD, all shareable data will be shared by adopting licenses that allow re-use of the data and of datasets in their entirety by other researchers/scholars and stakeholders of policy and market sectors.

The EC Open Access policy asks researchers to make available in Open Access their 'peer reviewed articles'. This is easiest to comply with when the researcher retains his/her copyright and only gives the publisher of the article a 'licence to publish'. In that case, the article can be deposited in a repository and made publicly accessible without further permission of the publisher. If the licence stipulates an embargo period, of course that must be respected.

Creative Commons licences are available in many countries for authors who wish to retain their copyright and provide their publisher with a licence.

## 8 Allocation of resources

GOLD will use standard tools and a free of charge research data repository. The costs of data management activities are limited to project management costs and will be covered by allocated resources in the project budget. Long-term preservation of the public data is ensured through a repository (ie. ZENODO). Other resources needed to support re-use of data after the project ends will be solved on a case-by-case basis.

The overall responsibility for data management lies with the project coordinator, Efthymia Alexopoulou from Cres. Supporting the coordinator is the project management team consisting of WP leaders and the exploitation team RE-CORD.

---

## 9 Rights to access and re-use of research data

Open Access to research refers to the right to access and re-use digital research data under the terms and conditions set out in the Grant Agreement. Openly accessible research data can typically be accessed, mined, exploited, reproduced, and disseminated free of charge for the user.

## 10 Data sets overview

The following table offers an overview of the expected datasets from the project and which will be collected in more detailed using the template in Annex I during the project. The table content will be updated according to the progress of the project.

| ID | WP | Task      | Project phase | Partner        | DATASET short title                              | STATUS                             |
|----|----|-----------|---------------|----------------|--|------------------------------------|
| 1  | 1  | 1.1       | M1-M8         | HUNAU          | Climatic data                                    | Available                          |
| 2  | 1  | 1.1 & 1.2 | M1-M48        | HUNAU          | Soil data  | 1.1= Available<br>1.2= in progress |
| 3  | 1  | 1.2       | M18-M48       | HUNAU          | Morphological data                               | Planned                            |
| 4  | 1  | 1.2       | M19-M48       | HUNAU          | Biomass yield data                               | Planned                            |
| 5  | 1  | T1.1      | M1-M9         | YNCREA - INRAE | Soil analysis / soil properties                  | Available                          |
| 6  | 1  | T1.1      | M1-M9         | YNCREA         | Climatic conditions                              | Available                          |
| 7  | 1  | T1.1      | M1-M9         | YNCREA         | Maps of total soil metal(loid)s                  | Available                          |
| 8  | 1  | T1.2      | M1-M36        | YNCREA- INRAE  | Plant parameters                                 | In progress                        |
| 9  | 1  | T1.2      | M1-M36        | YNCREA- INRAE  | Soil pore water                                  | In progress                        |
| 10 | 1  | T1.2      | M1-M36        | YNCREA- INRAE  | Analysis of plant parts                          | Planned                            |
| 11 | 1  | T1.3      | M4-M48        | YNCREA         | Pilot field trial                                | Planned                            |
| 12 | 1  | T1.4      | M7-M48        | YNCREA- INRAE  | Phytoremediation solutions                       | Planned                            |
| 13 | 1  | T1.1      | M4-7          | UNIBO          | Chiarini 2 site characterization and description | Available                          |
| 14 | 1  | T1.2      | M8-M12        | UNIBO          | Energy crops on potted contaminated soils        | In progress                        |

|    |   |       |         |       |   |             |
|----|---|-------|---------|-------|---|-------------|
| 15 | 1 | T1.3  | M12-M48 | UNIBO | Annual and perennial energy crops in contaminated site  | Planned     |
| 16 | 2 | 2.5.1 | M1-M32  | UDS   | High temperature autothermal pyrolysis and upgrading: Process configurations, conditions, products distribution and quality and contamination content in all derived product phases | In progress |
| 17 | 2 | 2.5.2 | M12-M36 | UDS   | High temperature autothermal pyrolysis and upgrading: CC - Process configurations, conditions, products distribution and quality  | In progress |
| 18 | 2 | 2.5.1 | M8-M48  | UDS   | High temperature autothermal pyrolysis and upgrading: FTS - Process conditions, catalysts, products distribution and quality  | In progress |

## Annex I – Data Collection Form

This Annex includes the template in which information about datasets will be collected for GOLD project.

GOLD will follow the Horizon 2020 Open Research Data guidelines and FAIR (Findable, Accessible, Interoperable, and Re-usable) principles. Since good research data management is not a goal in itself, rather a key leading to knowledge discovery and innovation, GOLD partners will ensure these data will be produced and archived according to the principles contained in the Data Management Plan (DMP). DMP will be established to support the data management life cycle for all data that will be duly collected, processed or generated by the project for guaranteeing its integration and reuse. The DMP will provide the identification elements and the descriptions of the data sets and it will include details regarding how the research data will be handled during the project and how they will be preserved after it is completed. It will specify which methodologies and standards will be used in the data creation and management and how and when the data will be shared and made open for re-use.

For each dataset the following elements will be defined:

- 1) data set name, identifier and description – including origin (if collected), scale and possible use and impact of the data;
- 2) metadata and standards;
- 3) details on data sharing, open access dissemination and licensing;
- 4) archiving and preservation. In case parts of the data cannot be openly shared, the DMP will provide justifications.

The DMP will be structured by ETA but co-developed in strict cooperation with each WP leader and integrated by a plenary contribution by all partners. Each partner going to publish an Open Access peer-reviewed scientific article is deemed responsible for the due management of related datasets and metadata.

**Please submit one form for each dataset you plan to collect in the project.**

Further updates of the DMP will be foreseen in a following project stage.

Submitted information will be kept confidential with access foreseen to the GOLD consortium and the European Commission only.

|                  |   |  |
|------------------|---|--|
| No.1             | In progress/<br>Not yet available<br><i>[please choose one]</i>   | <i>Dataset title: try to name it as follows:<br/>PROJECT ACRONYM: WP number: WP title aims: Task number: Task title:<br/>additional information specifying coverage and nature of data (if necessary):<br/>version number (in case of revisions and updates)</i> |
| Team in charge   | <i>Partner acronym<br/>[XXX]</i>  |  |
| Creator/s        | <i>Name/s of the person/s &amp; partner acronym who has the idea<br/>Name Surname [XXX]</i>                                       |  |
| Contributor/s    | <i>Name/s of the person/s &amp; partner who is/are contributing to this dataset<br/>Name Surname [XXX]</i>                        |  |
| Contact Person/s | <i>Technical referee (on behalf of the creator) for the dataset<br/>Name Surname [XXX, email address]</i>                         |  |
| Contents         | <i>Describe with all the info at disposal what the dataset is about, its purpose etc.<br/>This dataset contains data on XXXXX</i> |  |

|               |   |
|---------------|---|
|               | <p>Information includes XXXXX</p> <p>Each partner contributes by gathering data, implementing XXXXXX</p> <p>The dataset is made of X sheets, collecting information on: XXXXXX</p> <p>Data were already produced / are new XXXXXX</p> <p>Additional entities contributing to this dataset XXXXXX</p> <p>Potential users of the dataset include researchers, policymakers, etc.</p>  |
| Data format   | <p><i>Please indicate the type of data (e.g. if they are raw data, derived or secondary data, etc.).</i></p> <p><i>Please indicate also in which format they are provided. Examples could be:</i></p> <ul style="list-style-type: none"> <li>● Qualitative and quantitative tabular data (.csv, .rtf, .txt)</li> <li>● Qualitative textual data (.txt)</li> <li>● Statistical tabular data (.csv and STATA format: .dta)</li> <li>● .....</li> </ul>  |
| Data volume   | Final volume of data is expected to be XXX MB   |
| Accessibility | <p><i>Please try to indicate (if possible) if data will be made openly available or not; if so, indicate the type of license for their access. Examples/options could be:</i></p> <ul style="list-style-type: none"> <li>● Shareable data will be released as part of Deliverable X.X and other will be accessible under a CC BY 4.0 license</li> <li>● An ODC license</li> <li>● It is not defined yet whether the dataset will be made openly available as underlying data at the time of publication of research results or as other data</li> <li>● For information resulting from open access sources, such as Eurostat, there will be no protection. Data will be openly shared at the time of publication of results under a CC BY NC 4.0 license.</li> <li>● .....</li> </ul> |

## Annex II – Data Inventory Table

| Title | Type of data | Location |
|-------|--------------|----------|
|       |              |          |
|       |              |          |
|       |              |          |