



Deliverable D5.1



## FORGENIUS- Improving access to FOREst GENetic resources Information and services for end-USers

### *Deliverable D5.1*

## FORGENIUS Data Management Plan

**Planned delivery date (as in DoA): M6 30/06/2021**

**Actual submission date: 30/06/2021**

**Workpackage: WP5**

**Workpackage leader: GOZDIS**

**Deliverable leader: GOZDIS**

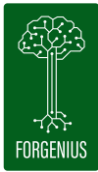
**Version: 1.0**

<b>Project funded by the European Commission within the Horizon 2020 Programme</b>	
<b>Dissemination Level: PU</b>	
<b>PU</b> Public	
<b>CI</b> Classified, as referred to Commission Decision 2001/844/EC	
<b>CO</b> Confidential, only for members of the consortium (including the Commission Services)	

Research and Innovation action: GA no. 862221

Start date of the project: January 1<sup>st</sup>, 2021





## Deliverable D5.1

### **Authors**

EFI, GIS, INRAE

### **Reference**

D5.1: FORGENIUS Data Management Plan, H2020 project no. 862221 SFS-28-2019-European Commission, 2021

### **Executive summary**

The main aim of the FORGENIUS Data Management Plan (DMP) is to provide details on the principles and procedures associated with data management. This document lays out how Horizon 2020 open access data policies are implemented in the FORGENIUS project. Open access to publications generated by the FORGENIUS consortium is mandatory, as well as to the data associated with these publications. To comply with this policy, both the papers and associated datasets will be stored in public and credible repositories. Additionally, FORGENIUS will upgrade the European Information System on Forest Genetic Resources (EUFGIS), where relevant data generated by the project will be stored and made available during the project and for the long term. Annex 2 describes the DMP template that will be used to generate the DMPs associated with each FORGENIUS dataset.



## Acronyms

ASCII	American Standard Code for Information Interchange
CERN	Conseil Européen pour la Recherche Nucléaire, or European Council for Nuclear Research
CNR	Consiglio Nazionale delle Ricerche- Italian National Research Council
DCC	Digital Curation Centre
DMP	Data Management Plan
DOI	Digital object identifier
EFI	European Forest Institute
EUFGIS	The European Information System on Forest Genetic Resources
FAIR	Findable, accessible, interoperable and re-usable
FGR	forest genetic resources
FORGENIUS	Improving access to FOReSt GENetic resources Information and services for end-USers
GCU	Genetic conservation unit
H2020	Horizon 2020 ,
INIA	Instituto Nacional de Investigaciones Agronómicas – Spanish National Institute for Agricultural and Food Research and Technology
INRAE	Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement - French National Research Institute for Agriculture, Food and Environment
CREAF	Centro de Investigación Ecológica y Aplicaciones Forestales - Spanish Ecological and Forestry Applications Research Centre
JRC	Joint Research Centre - European Commission
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized difference vegetation index
NIRS	Near infra-red spectra
OA	Open access
UKDS	UK Data Service
UMR	Philipps-Universität Marburg (Germany)
VCF	Variant Call Format



## TABLE OF CONTENTS

1.	Introduction.....	5
1.1	Project summary.....	5
2.	Open access to research data.....	7
2.1	Context .....	7
2.2	Data Types and Data Sharing.....	9
3.	Long-term Repositories .....	11
4.	Project Repository .....	13
5.	The European Information System on Forest Genetic Resources (EUFGIS) .....	13
6.	Annexes .....	14
	Annex 1: Definitions .....	14
	Annex 2: FORGENIUS DMP template .....	15
	Annex 3. Creative Commons license .....	18
	Annex 4: Response to questions in H2020 Data management update template.....	19



## 1. Introduction

The objectives of this document are to:

- Identify the different types of data (origin and authorship, see Annex 1 for definition of terms) that will be collected during the project
- Define the different categories of users for these data during FORGENIUS and after it is finished
- Provide documents and recommendations for data management: a Data Management Plan template (Annex 2); and data licences adapted to specific Intellectual Properties using the Creative Commons licences (Annex 3)
- Provide support for producing and managing discovery metadata, according to metadata standards that comply with international requirements and constraints, including developing a metadata management system.

This document follows FAIR (findable, accessible, interoperable and re-usable) data management rules. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution. This document is not intended as a strict technical implementation of the FAIR principles. It is rather inspired by FAIR as a general concept. For more information about FAIR, see its data principles<sup>1</sup> and the *FAIR Guiding Principles for scientific data management and stewardship*<sup>2</sup>.

This document will be updated again during the duration of the project after significant changes arise, such as new data and new data types; progress in upgrading EUFGIS; changes in consortium policies, and changes in consortium composition and external factors.

### 1.1 Project summary

The European Commission, Research Executive Agency, is – under the Horizon 2020 framework programme – funding an action entitled “Improving access to FORest GENetic resources Information and services for end-USers” — ‘FORGENIUS’.

FORGENIUS is a Research and Innovation Action which aims to develop methods and tools providing greater insight into the characteristics and the value of all relevant forest genetic resources (FGR) presently conserved in 35 European countries, and linked

<sup>1</sup> <https://www.force11.org/group/fairgroup/fairprinciples>

<sup>2</sup> Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PA, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016 Mar 15;3:160018. doi: 10.1038/sdata.2016.18. Erratum in: *Sci Data*. 2019 Mar 19;6(1):6. PMID: 26978244; PMCID: PMC4792175.



## Deliverable D5.1

through the European Information System on Forest Genetic Resources (EUFGIS) (<http://portal.eufgis.eu>). FORGENIUS will create novel services for users within and outside the conservation communities and increase and improve data quantity and quality in EUFGIS (see Chapter 5). The services created by FORGENIUS will also allow end-users (the National Focal Points in the 35 countries) to characterise prospective new genetic conservation units.

The project will use state-of-the-art technology and knowledge in plant and evolutionary biology, ecology, remote-sensing, genomics, genetics, modelling, and forestry science. It will produce a platform allowing predictions of the fate of European conserved genetic resources in the short-, medium-, and long term. Such a tool will be critical for developing new, adaptive strategies to preserve FGR at the continental level.

The overall goal of FORGENIUS is to produce new high-throughput quantitative assessments of conserved FGR and make these accessible to end-users by developing general standards, tools and services for a better characterization and management of the entire network of genetic conservation units (GCUs)<sup>3</sup>. To do this, FORGENIUS will:

- Assess genetic and phenotypic diversity of a pilot set of GCUs, as well as their resilience, under climate change
- Apply the novel data to underpin management decisions that promote GCU resilience and adaptability
- Characterise genetic resources to identify high-quality germplasm for use in breeding and plantations
- Create innovative data query, analysis and synthesis services for end-users within and beyond the conservation communities.

Tree species studied are: *Abies alba*, *Abies nebrodensis*, *Betula pendula*, *Castanea sativa*, *Fagus sylvatica*, *Fraxinus excelsior*, *Malus sylvestris*, *Picea abies*, *Pinus halepensis*, *Pinus nigra*, *Pinus pinaster*, *Pinus pinea*, *Pinus sylvestris*, *Populus alba*, *Populus nigra*, *Prunus avium*, *Quercus ilex*, *Quercus petraea*, *Quercus pubescens*, *Quercus robur*, *Quercus suber*, *Sorbus torminalis*, *Taxus baccata*.

<sup>3</sup> GCUs are networks of forest stands or areas harbouring tree populations which have adapted to specific environmental conditions or have distinct characteristics. Such stands are typically located in forests managed for multiple uses, protected areas or seed stands. For more information, please visit <http://portal.eufgis.org/genetic-conservation-units>



## 2. Open access to research data

### 2.1 Context

Open access to research data refers to the right to access and reuse digital research data under the terms and conditions set out in the Grant Agreement. Research data refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings, and images amongst others. The focus is on research data that is available in digital form. Users can normally access, data-mine, exploit, reproduce and disseminate openly-accessible research data free of charge. Open access to data used in scientific publications is mandatory, while for other data types it is recommended. Data used in scientific publication must also be uploaded to a non-commercial institutional repository and associated to a Permanent Unique Identifier (e.g. DOI). Guidance and services for the publication of FAIR datasets can be found in the OpenAIRE portal ([www.openaire.eu](http://www.openaire.eu)). Many research institutes, universities or research communities have developed such data repositories that can be found in registries like [re3data.org](http://re3data.org) or [fairsharing.org](http://fairsharing.org). See paragraphs below for recommendations.

All FORGENIUS research related activities that collect or generate data should provide a Data Management Plan (DMP) specific to that data. The H2020 DMP template (Annex 2) should be preferably used by FORGENIUS partners. Each dataset must be associated with its relevant metadata, including its authors, title and description plus any information on how the data will be shared, stored (including backup) and preserved in the long term (archived). If a dataset is planned to be made public, in particular in association with a scientific paper, it must have a DOI<sup>4</sup> (Digital Object Identifier – see also section on repositories) to follow open-access recommendations.

It is recommended that FORGENIUS partners store their data (or at least the metadata) on the FORGENIUS data repository, managed by the European Forest Institute (EFI) in its servers. Data will be immediately accessible by all partners, without any embargo. Studies based on specific datasets must be proposed as co-research activities with the partner(s) owning the intellectual property rights.

Data collected or generated from the different FORGENIUS activities are strategic resources for the whole FORGENIUS consortium. They include only original data that will be generated during the project. These are databases, and sampling results and surveys that will lead to publications (reports) for which they will be analysed, and should include metadata, i.e. the information describing the managed research data. Specific groups of data produced in the project are presented in chapter **Data Types and Data Sharing** below.

<sup>4</sup> [https://www.doi.org/overview/DOI\\_article\\_ELIS3.pdf](https://www.doi.org/overview/DOI_article_ELIS3.pdf)



## Deliverable D5.1

This document is not concerned with any data pre-existing the FORGENIUS project. These data are part of the “background” contributions supplied by the FORGENIUS partners and are identified in the Consortium Agreement document. This document is also not concerned with data originating from third party producers. These data are managed according to the rules defined by these third parties. The FORGENIUS data policy is based on the full sharing of data among partners of the project, and on open data access for external users, except when this is not possible (see list below).

FORGENIUS encourages the publication of data papers making the data produced by FORGENIUS accessible to the community of researchers or stakeholders interested in managing genetic resources. FORGENIUS encourages authors of publications using FORGENIUS data to clearly identify and acknowledge such data, either by citing the relevant data papers or by linking to the data DOI. The open access data must be attributed with an appropriate Creative Commons licence (Annex 3), depending on the type of dataset and on the type of uses.

Metadata should be set according to the [DataCite Metadata Schema \(http://schema.datacite.org/\)](http://schema.datacite.org/), with properties listed in Table 2.

Table 2: DataCite metadata properties

ID	Property	Status of property
1	Identifier (with type attribute)	Mandatory
2	Creator (with name identifier)	Mandatory
3	Title (with optional type attribute)	Mandatory
4	Publisher	Mandatory
5	Publication Year	Mandatory
6	Subject (with schema attribute)	Optional
7	Contributor (with type and name)	Optional
8	Date (with type attribute)	Optional
9	Language	Optional
10	Resource Type (with description attribute)	Optional
11	Alternate Identifier (with type attribute)	Optional
12	Related Identifier (with type and relation type attributes)	Optional
13	Size	Optional
14	Format	Optional
15	Version	Optional
16	Rights	Optional
17	Description (with type attribute)	Optional
0.1	LastMetadataUpdate	Administrative
0.2	MetadataVersionNumber	Administrative





## 2.2 Data Types and Data Sharing

FORGENIUS Project will produce three types of data which will be stored accordingly:

	Data types	Long-term Storage
a.	Raw Data	Project partners
b.	Curated Data	EUFGIS
c.	Data Summaries	EUFGIS

### a. Raw data

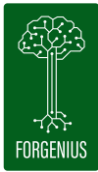
Raw data is data in its native format, in general as produced by an instrument or machine; no transformations have been applied to the raw data. No overwriting is allowed. This type of data is not ready to use, and it requires prior transformation with algorithms / computer programs that are specific in terms of appropriate and relevant use. These data will be stored by the project partners (see acronym list for full partner names) who are responsible for their production:

- INIA: genetic data (WP3)
- INIA/INRAE: NIRS data (near infra-red spectra acquired for drill dust, cores and branch samples, WP3).
- CNR: genetic data (WP4)
- UMR: data from unmanned aerial vehicles (WP3)
- INRAE/CREAF: plant physiology data (WP1/WP2/WP3)
- INRAE/JRC: remote sensing data

Examples of raw data:

- DNA sequence data (“sequencing reads”, typically in FASTQ format) obtained from DNA sequencers
- Shoot hydraulic conductance values obtained from ‘cavitron’ equipment (table format)
- Raw images on different wavelength bands from different satellites: MODIS, SENTINEL2
- Spectra collected from benchtop and portable NIRS equipment are generated as proprietary format files

These data will be used as part of the project to create curated data and data summaries. End-users will not have access to this data.



## Deliverable D5.1

### **b. Curated data**

Curated data is cleaned raw data that is converted into consumable records that can be stored in files or tables. Curated data are uniform in the way files are stored in terms of encoding, format, data types, and content (i.e., strings). These data will be stored in EUFGIS.

Examples of curated data:

- Genotype tables in VCF format, as obtained after the processes of “read mapping” and “variant calling” from “sequencing reads”
- Parameters (e.g. slope and intercept) of individual-sample xylem vulnerability curves, as obtained from cavitron data
- Seasonal dynamics of indices such as NDVI on the GCU since 2000 obtained from MODIS
- NIRS spectra tables converted from proprietary format to text or ASCII delimited files

These data will be used in the project to produce data summaries. Access to these data will be granted to end users in accordance with the EUFGIS’ Data Sharing Agreement.

### **c. Data summaries**

Data summaries are synthetic values, that are produced by the application of a mathematical function to data from multiple samples (e. g. means, variances of curated data, indices of genetic diversity, amongst others). Data summaries can describe GCUs using indices developed in the project. It will be possible to use these data summaries to analyse patterns or identify correlations. These data will be stored in EUFGIS.

Examples of exploration data:

- Allele frequencies for DNA variants; gene diversity
- Mean and variance of xylem sensitivity to embolism
- Yearly averaged values of NDVI, temperature and land cover obtained from GCUs using remote-sensing, such as in MODIS, LANDSAT or SENTINEL2.
- Predicted ecophysiology parameters from NIRS
- Microdensity data from X-ray analysis

These data will be used in the project to produce outputs and will be available via the EUFGIS web interface. Access to these data will be unrestricted.



### 3. Long-term Repositories

Grant Beneficiaries must deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications – together with the data behind it. Even if the scientists have chosen ‘gold’ open access, they are still obliged to deposit the publication in a repository. This must be done at the latest 6 months after publishing. Publications must be in a format that can be used and understood by a computer, i.e. 'Machine-readable electronic copy'. They must be stored in text file formats that are either standardised or otherwise publicly known so that anyone can develop new tools for working with the documents. A 'repository' for scientific publications is an online archive. Institutional, subject-based and centralised repositories are all acceptable choices. Repositories that claim rights over deposited publications and preclude access are not acceptable. If authors are unsure which repository to use, they should consult [OpenAIRE](#) pages first. Most of research institutes and faculties in Europe have their own repository, and they can be browsed at: <https://www.openaire.eu/participate/deposit-publications-data>.

If an organization does not have a repository and does not know which one to use – we recommend that they use [Zenodo](#), CERN's repository. In Zenodo, metadata is stored internally in JSON-format according to a defined [JSON schema](#). Metadata is exported in several standard formats such as MARCXML, Dublin Core, and DataCite Metadata Schema (according to the [OpenAIRE Guidelines](#)). Within this and other repositories, anyone can store both publications and data, and also supplementary material, such as posters, presentations, images, etc. If the material planned for uploading to a repository does not have a DOI, in most repositories it is automatically generated with the upload. The upload page hosts all mandatory metadata properties according to the DataCite Metadata Schema – so no one needs to prepare a separate metadata file in order to upload it to the repository (although for the duration of the project, project partners can upload data in the common repository see chapter 4 below). The repositories also allow for specifying access rights, for example: (1) open access; (2) embargoed access (specifying the date from which it is available, in accordance with publishing agreement); (3) restricted access (with user-specified conditions), and (4) closed access. If the open access option is selected, the Creative Commons licence will also need to be selected (see Annex 3) anyhow, the project consortium urges for open access for FORGENIUS data and papers.

Grant Beneficiaries must ensure that the metadata associated with deposition in the repository include the acknowledgement of the EU funding program according to the following standard format:

- terms [*"European Union (EU)" & "Horizon 2020"*][*" Sustainable Food Security 2018-2020 Genetic resources and pre-breeding communities "*]
- FORGENIUS - Improving access to FORest GENetic resources Information and services for end-USers, grant agreement No 862221.



## Deliverable D5.1

- publication date, the length of the embargo period (if applicable) and a persistent identifier.

This is very important to ensure that Horizon 2020 can be properly monitored, statistics produced, and the program's impact assessed.

To monitor any embargo periods, the publication date and embargo period must be provided. The persistent identifier (for example a Digital Object Identifier) identifies the publication. It enables a link to be provided to an authoritative version of the publication.

- In all cases, the Commission encourages authors to retain their copyright and grant adequate licences to publishers. [Creative Commons](#) offers useful licensing solutions (See Annex 3). This type of licence is a good legal tool for providing open access in its broadest sense.

Where possible, contributors should also be uniquely identifiable, and data uniquely attributable, through identifiers which are persistent, non-proprietary, open and interoperable (e.g. through leveraging existing sustainable initiatives such as [ORCID](#) for contributor identifiers and [DataCite](#) for data identifiers).

Example of a short explanation on what to do with a published paper and the data behind it:

1. Prepare the PDF of authors' accepted manuscript with a DOI of the original publication, its full reference and other information when needed. State FORGENIUS as the funding source. Authors can use the following wording: *This research was conducted as part of the H2020 Project FORGENIUS (Improving access to FORest GENetic resources Information and services for end-USers) project (grant agreement No 862221).*
2. Prepare the Data Management Plan PDF file for the data behind the publication (explanation of the dataset and rules on how it should be used – see Annex 2 and Table 1)
3. Upload the authors' accepted manuscript, data and the data management file to a non-commercial institutional repository (recommendation is Zenodo if the author's institution does not have its own repository). Set the embargo period and the Creative Commons license according to the rules of the journal.
4. Upload all the files to FORGENIUS Intranet – <https://sites.inra.fr/site/forgenius/SitePages/Home.aspx>.



## 4. Project Repository

In order to simplify transfer of raw data between project partners and to ensure an easy and standardised repository for raw data during the project's duration, EFI will develop a temporary repository on its servers. This repository will be made available by month 12 and all project partners will be provided credentials to access the repository.

## 5. The European Information System on Forest Genetic Resources (EUFGIS)

The FORGENIUS project will upgrade the European Information System on Forest Genetic Resources (EUFGIS- <http://portal.eufgis.org>), the only transnational information system on FGR in Europe. At the start of the project, it contained information on more than 3500 Genetic Conservation Units (GCUs) of 111 tree species from 35 European countries.

Since the establishment of the EUFGIS in 2010, European countries have started to follow the “pan-European minimum requirements for dynamic genetic conservation units of forest trees”

([http://portal.eufgis.org/fileadmin/templates/eufgis.org/documents/EUFGIS\\_Minimum\\_requirements.pdf](http://portal.eufgis.org/fileadmin/templates/eufgis.org/documents/EUFGIS_Minimum_requirements.pdf)). The minimum requirements are based on the concept of dynamic conservation of genetic diversity, which emphasises the maintenance of evolutionary processes within tree populations to safeguard their potential for continuous adaptation. The units entered into the EUFGIS database have a designated status as genetic conservation areas of forest trees at national level. The minimum requirements also specify a minimum size of a unit, depending on tree species and conservation objectives. The EUFGIS information system is regularly maintained by National Focal Points in European countries (<http://portal.eufgis.org/data-providers>).

Up to date, all data entered in EUFGIS have been characterised by limited climatic and environmental parameters. The FORGENIUS project will link data from existing Europe-wide databases on climate and environment, allowing a very precise and continuously improving characterisation of the data. FORGENIUS will also characterise GCUs by remote sensing data. In addition, FORGENIUS will genetically and phenotypically characterise GCUs of 23 species.

All these data that characterise the GCUs will be made available in EUFGIS, which is maintained by the European Forest Genetic Resources Programme (EUFORGEN).



## 6. Annexes

### Annex 1: Definitions

**Background:** Information, Know How, DataBase, Software or Material which are held by a Party prior to its accession to the Consortium Agreement or developed outside the scope of this Consortium Agreement, as well as copyrights or other intellectual property rights pertaining to such Information, Know How, DataBase, Software or Material, for which the application has been filed before its accession to this Consortium Agreement, or up to the termination of the Consortium Agreement, and which is needed for carrying out the Project or for using the Foreground.

**Creative Commons (CC) Licenses:** Legal tools that are globally recognized and allow users in a simple and standardized way to describe the conditions of access and reuse of their work.

**Data:** All scientific information produced by the project

**Open Data:** Data freely available/accessible to any User. They are made available under an appropriate license that takes into account intellectual property issues, such as copyright, acknowledgement/attribution, identification, and specific re-use conditions.

**Intellectual Property Rights:** the protections granted to the legal owners of intellectual creations in all fields of human endeavour, including the industrial, scientific, literary and artistic fields

**User:** A person who uses or makes use of the data produced by the project, directly or indirectly

**Metadata:** Information about data provenance, authorship, description, quality, processing (raw data, elaborated data, derivative product), and collection/generation context, which supports interoperability across disciplines. FORGENIUS metadata shall meet or exceed applicable national or European requirements.



## Annex 2: FORGENIUS DMP template

The template is based on H2020 and DCC<sup>5</sup> templates (<http://www.dcc.ac.uk/resources/how-guides/develop-data-plan>).

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the applicants with regard to all datasets that will be generated by the project. The DMP is not a fixed document; it evolves during the lifespan of the project. The DMP should address the points below **on a dataset-by-dataset basis** and should reflect the current status of reflection within the consortium about the data that will be produced.

FORGENIUS data should be discoverable, accessible, assessable and intelligible, useable beyond the original purpose for which they were collected and, finally, interoperable to specific quality standards.

Each beneficiary should keep its own record of all the produced data. They can in addition or alternatively be store in the repository made available by EFI.

**For each data set**, the following five main items should be addressed and specified:

### 1- Data set reference and name

Identifier for the data set to be produced.

### 2- Data set description

Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and re-use.

Questions to consider on data set description:

- What data will you create?

Guidance:

- Give a brief description of the data that will be created, noting its content and coverage

### 3- Standards and metadata

Reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.

Questions to consider on standards and metadata:

<sup>5</sup> The Digital Curation Centre (DCC) is a world-leading centre of expertise in digital information curation with a focus on building capacity, capability and skills for research data management.



## Deliverable D5.1

- How will you capture/create the metadata?
- Can any of this information be created automatically?
- What metadata standards will you use and why?

Guidance:

- Metadata should be created to describe the data and aid its discovery. Consider how you will capture this information and where it will be recorded, e.g. in a database with links to each item, in a 'readme' text file, in file headers etc.
- Researchers are strongly encouraged to use community standards to describe and structure data, where these are in place. The DCC offers a catalogue of disciplinary metadata standards.

### **4- Data sharing**

Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.). In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

Questions to consider on methods for data sharing:

- How will you make the data available to others?
- With whom will you share the data, and under what conditions?

Guidance:

- Consider where, how, and to whom the data should be made available. Will the data be shared via a data repository, data requests handled directly, or another mechanism used?
- The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. Mention earlier examples to show a track-record of effective data sharing.

### **5- Short- and long-term preservation (including backup)**

Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its estimated end volume, what the associated costs are and how these are planned to be covered.

Questions to consider on short and long-term preservation:





## Deliverable D5.1

- What is the long-term preservation plan for the dataset? e.g. deposit in a data repository
- Will additional resources be needed to prepare data for deposit or meet charges from data repositories?

### Guidance:

- Researchers should consider/ make provision for how datasets with long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving.
- If authors/ data managers do not propose to use an established repository, the data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the grant.








### Questions to consider on Resourcing:

- What additional resources are needed to deliver your plan?
- Is additional specialist expertise (or training for existing staff) required?
- Do you have sufficient storage and equipment, or do you need to cost in more?
- Will charges be applied by data repositories?
- Have you costed in time and effort to prepare the data for sharing / preservation?

### Guidance:

- Carefully consider any resources needed to deliver the plan. Where dedicated resources are needed, these should be outlined and justified. Outline any relevant technical expertise, support and training that is likely to be required and how it will be acquired. Provide details and justification for any hardware or software which needs to be purchased or additional storage and backup costs that may be charged by IT services.
- Funding should be included to cover any charges applied by data repositories, for example to handle data of exceptional size or complexity. Also remember to cost in time and effort to prepare data for deposit and ensure it is adequately documented to enable re-use. If you are not depositing in a data repository, ensure you have appropriate resources and systems in place to share and preserve the data.
- See UKDS guidance on costing data management (<https://www.ukdataservice.ac.uk/manage-data/plan/costing>).

### Annex 3. Creative Commons license

No	Name	Mark and Code	Summary	Links
0	No Rights Reserved	 PUBLIC DOMAIN	This license puts a work in to the public domain by waiving all rights to the work worldwide under copyright law, including all related and neighbouring rights, to the extent allowed by law. It is possible to modify, distribute and perform the work, even for commercial purposes, all without asking permission. No attribution is required	<a href="#">View description</a>
1	Attribution	 CC BY	This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.	<a href="#">View License</a> <a href="#">Legal Code</a>
2	Attribution-NoDerivs	 CC BY-ND	This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.	<a href="#">View License</a> <a href="#">Legal Code</a>
3	Attribution-NonCommercial-ShareAlike	 CC BY-NC-SA	This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.	<a href="#">View License</a> <a href="#">Legal Code</a>
4	Attribution-ShareAlike	 CC BY-SA	This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects	<a href="#">View License</a> <a href="#">Legal Code</a>
5	Attribution-NonCommercial	 CC BY-NC	This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.	<a href="#">View License</a> <a href="#">Legal Code</a>
6	Attribution-NonCommercial-NoDerivs	 CC BY-NC-ND	This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.	<a href="#">View License</a> <a href="#">Legal Code</a>



## **Annex 4: Response to questions in H2020 Data management update template**

### 1. Data Summary

The European Commission, Research Executive Agency, is – under the Horizon 2020 framework programme – funding an action entitled ‘Improving access to FORest GENetic resources Information and services for end-USers’ — ‘FORGENIUS’.

FORGENIUS is a Research and Innovation Action which aims to develop methods and tools providing greater insight into the characteristics and the value of all forest-relevant forest genetic resources (FGR) collections presently existing in 35 European countries, and linked through the EUFGIS Information System (<http://portal.eufgis.eu>). FORGENIUS will create novel services for users within and outside the conservation communities and significantly increase and improve data quantity and quality in the European FGR information system that describes all accessions. The services created by FORGENIUS will also allow end-users to characterise prospective new genetic conservation units.

### 2. FAIR data: Making data findable, including provisions for metadata

In this project there is no data produced and/or used in the project to be made discoverable with metadata, identifiable and locatable by means of a standard identification mechanism.

### 3. Data security

Provisions that are in place for data security regarding project information.

EFI has general technical and organisational measures in order to ensure compliance with these rules, for data stored in the project’s repository and in EUFGIS; INRAe has general technical and organisational measures for data, as the project coordinator for data not stored in EUFGIS. Project Partners have responsibilities on technical and organisational measures in order to ensure compliance with these rules, for raw data stored on their servers.

These measures especially focus on:

- prevention of unauthorised persons from gaining access to automated systems or filing systems
- only allowing authorised users of an automated system or filing system to access the personal data to which their access right refers
- recording which personal data has been communicated, when and to whom
- preventing that, during communication of personal data and transport of storage media, the data can be read, copied or erased without authorisation



## Deliverable D5.1

The data safely stored in certified repositories for long term preservation and curation.

### 4. Ethical aspects

There are no ethical or legal issues that can have an impact on data sharing. These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

All these issues are addressed in the FORGENIUS two ethical deliverables in WP9:

D 9.1 (H - Requirement No. 1) Equality of treatment & Informed consent

D 9.2 (POPD - Requirement No. 2) POPD - Data protection

### 5. Other issues

The project does not make use of other national/funder/sectorial/departmental procedures for data management.