

GenTree

Optimizing the management and sustainable use of forest genetic resources in Europe

Deliverable D7.3

GenTree data management plan

Planned delivery date (as in DoA): M3

Actual submission date: 08/12/2016, M10

Workpackage: WP7

Workpackage leader: INRA

Deliverable leader: INRA

Version: 1.0

This deliverable is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676876.	
Dissemination Level	
PU Public	PU
CI Classified, as referred to Commission Decision 2001/844/EC	
CO Confidential, only for members of the consortium (including the Commission Services)	

Research and Innovation action: GA no. 676876

Start date of the project: March 1st, 2016



TABLE OF CONTENTS

Administrative details of the project	3
1- Introduction and purposes	3
2- Data type and access	4
3- Data Management Plan	5
4- Data sharing agreement for GenTree partners	5
5- Data identification/citation	6
6- Metadata standards and tools for discovery	6
Annex 1: Definitions	7
Annex 2: GenTree DMP template	8
Annex 3: GenTree Data Sharing Agreement	11
Annex 4: Creative Commons Licenses	13

Data Management Plan for H2020 Research and Innovation action 676876 (DMPGenTree)

Administrative details of the project

Project Name: Optimising the management and sustainable use of forest genetic resources in Europe (GenTree) - **DMP title:** DMPGenTree

Start date: 1st March 2016 - **Duration:** 48 months

Coordinator: Dr. Bruno Fady, INRA, France (bruno.fady@inra.fr)

Funder: European Commission (Horizon 2020)

1- Introduction and purposes

The goal of GenTree is to provide the European forestry sector with better knowledge, methods and tools for optimising the management and sustainable use of forest genetic resources (FGR) in Europe in the context of climate change and continuously evolving demands for forest products and services. To reach its goal, GenTree will make scientific, technological and implementation breakthroughs in: (i) designing innovative strategies for dynamic conservation of FGR in European forests, (ii) broadening the range of FGR used by European breeding programmes, and (iii) preparing new forest management scenarios and policy frameworks fully integrating genetic conservation and breeding aspects, to adapt forests and forestry to changing environmental conditions and societal demands. GenTree focuses on economically and ecologically important tree species in Europe, growing in a wide range of habitats and covering different societal uses and values.

The major outputs of GenTree will include: (i) much needed new scientific knowledge on phenotypic and genotypic diversity across environmental gradients in Europe, (ii) improved genotyping and phenotyping monitoring tools for practitioners, (iii) updated and refined data for information systems of in-situ and ex-situ FGR collections, (iv) innovative strategies for conservation, breeding and exchanging and using diversified forest reproductive material, (v) novel outreach and science-policy support tools to better integrate FGR concerns into forest management and better implement relevant international commitments in Europe. GenTree will improve the status and use of European in-situ and ex-situ FGR collections, support acquisition, conservation, characterisation, evaluation and use of relevant FGR in breeding and forestry practice and policy, will seek to harmonise, rationalise and improve management of existing collections and databases, and will strengthen the EU strategy for cooperation on FGR research and innovation. Tree species studied are:

- Conifers: *Abies alba*, *Picea abies*, *Pinus cembra*, *Pinus halepensis*, *Pinus nigra*, *Pinus pinaster*, *Pinus sylvestris*, *Taxus baccata*.
- Broadleaves: *Betula pendula*, *Fagus sylvatica*, *Populus nigra*, *Quercus petraea*.

Data collected or generated from the different GenTree research activities **are strategic resources** for the whole GenTree consortium. They are also valuable resources for a larger community of researchers or stakeholders interested in forest tree genetics and the management and sustainable use of forest genetic resources in Europe. Research data in

GenTree are numbers, texts, images and sounds generated by the partners during the project and **needed to validate** the results presented in scientific publications and other communication documents. They include metadata, i.e. the information describing the managed research data.

This document does not concern data preexisting the GenTree project. These data are part of the “background” contributions supplied by the GenTree partners and are identified in the Consortium Agreement document. This document does not concern either data originated from third party producers. These data are managed according to the rules defined by these third parties.

With the objective of **stimulating the use and re-use of data** and to contribute to the development of the “Open Science” strategy of the European Union through **data sharing**, the GenTree data policy framework provides recommendations and rules for data management and accessibility. It is based on existing documents published in the scientific literature, such as Michener (2015, "Ecological data sharing" Ecological Informatics) and the guidelines produced by the European Commission (2016, Guidelines on Data Management in Horizon 2020).

The **general purpose** of this document is:

- a) to identify the different types of data (origin and authorship, see annex 1 for definition of terms) that will be collected during GenTree and define the different categories of users for these data during GenTree and after it is finished.
- b) to provide documents and recommendations for data management and publication:
 - a Data Management Plan template (Annex 2);
 - recommendations for data sharing (Annex 3);
 - data licenses adapted to specific Intellectual Properties (Annex 4);
- c) to provide support for the production and management of discovery metadata:
 - metadata standards compliant with international requirements and constraints;
 - a metadata management system.

This document only focuses on data, including output data from modeling activities. Models and modeling platforms involved in the GenTree project are out of its scope. Intellectual property and sharing rules related to these components are addressed in the Consortium Agreement. Plant material and DNA are also outside the scope of this document. Their exchange and transfer will be addressed in a separate Material Transfer Agreement document.

2- Data type and access

The GenTree 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 not possible.

The data access policy is described in the Table below where different Creative Commons licenses (Annex 4) are assigned depending on the type of dataset and on the type of uses.

Type of data set	Type of uses (and users)			
	<i>GenTree research partner</i> <i>GenTree partners during the project</i>	<i>Non commercial 1</i> <i>Public research and education & GenTree partners after the end of the project</i>	<i>Non commercial 2</i> <i>Other users for non-commercial use</i>	<i>Commercial use</i> <i>Any user</i>
Data set type 1 <i>Metadata describing the genetic resources (sites, plots and trees)</i>	<i>No license</i> <i>Sharing (with authorship) among partners as described in the Data Sharing Agreement</i>			
Data set type 2 <i>Genetic and genomic data (DNA sequences, genotypic data, etc)</i>	<i>No license</i> <i>Sharing (with authorship) among partners as described in the Data Sharing Agreement</i>			
Data set type 3 <i>Phenotypic data (traits measured in situ and in the lab, related to phenotypic descriptions)</i>	<i>No license</i> <i>Sharing (with authorship) among partners as described in the Data Sharing Agreement</i>			

3- Data Management Plan

All GenTree research activities that collect or generate data should provide a Data Management Plan (DMP). Although there is no unique rule to define at which scale the DMP should be elaborated, it must correspond to a coherent data set relative to their management. The H2020 DMP template (Annex 2) should be preferably used by GenTree partners. For each data set, the following information needs to be provided: Data set reference and name, data set description, standards and metadata used, how data will be shared, how data will be archived and preserved (including storage and backup).

4- Data sharing agreement for GenTree partners

A data sharing agreement (Annex 3) has to be signed by all GenTree partners in order to ensure the full sharing of data among partners. Data will be immediately accessible for all

partners, without any embargo. Studies based on specific data sets must be proposed as co-research activities to the partner(s) owning the intellectual property rights.

5- Data identification/citation

GenTree data will be stored in GnpIS (<https://urgi.versailles.inra.fr/Tools/GnpIS>), a dedicated multispecies integrative Information System which ensures confidential treatment, secured archiving, browsing, visualizing and use and re-use of data. GnpIS is a licensed information system that will be used as is, without further development from its current state, by GenTree partners. GnpIS cannot be installed at GenTree partner laboratories.

GenTree recommends that data sets should be identified by a DOI. GnpIS makes it possible to generate DOI for data sets. Data and their DOI will remain accessible to partners for a minimum of 10 years after they are stored in GnpIS.

GenTree encourages the publication of data papers making the data produced by GenTree accessible to the community of researchers or stakeholders interested in the management and sustainable use of forest genetic resources in Europe. GenTree encourages that publications using GenTree data clearly identify and acknowledge such data and their Information System, either by citing the relevant data papers or by linking to the proper DOI in the GenTree Information System.

6- Metadata standards and tools for discovery

All GenTree datasets should be described by standardized metadata for discovery purposes. GenTree metadata will be stored and managed using the same dedicated Information System as for data, GnpIS (<https://urgi.versailles.inra.fr/Tools/GnpIS>). Metadata will follow the ISO19115/19139 standards and will be compliant with the EU INSPIRE directive. GenTree metadata will be fully and freely accessible.

Each Partner shall supply GnpIS, the GenTree Information System, with its metadata as soon as data are generated. Each Partner shall also supply the GenTree Information System with Metadata from third parties, according to the rules defined by such third party.

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 grant to 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 endeavor, 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, description, quality, processing (raw data, elaborated data, derivative product), and collection/generation context, which supports interoperability across disciplines. GenTree metadata shall meet or exceed applicable national or European requirements.

Annex 2: GenTree DMP template

Templates are based on H2020 and DCC 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 the datasets that will be generated by the project. The DMP is not a fixed document, but 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.

GenTree 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. In GenTree, there are 3 main categories of data: site description, DNA sequences and phenotypic traits.

For each data set, the following 5 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:

- 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 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 you share data via a data repository, handle data requests directly or use another mechanism?
- 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.

Questions to consider on restrictions for sharing:

- Are any restrictions on data sharing required? e.g. limits on who can use the data, when and for what purpose.
- What restrictions are needed and why?
- What action will you take to overcome or minimise restrictions?

Guidance:

- Outline any expected difficulties in data sharing, along with causes and possible measures to overcome these. Restrictions to data sharing may be due to participant confidentiality, consent agreements or IPR. Strategies to limit restrictions may include: anonymising or aggregating data; gaining participant consent for data sharing; gaining copyright permissions; and agreeing a limited embargo period.

Questions to consider on data repository:

- Where (i.e. in which repository) will the data be deposited?

Guidance:

- Most research funders recommend the use of established data repositories, community databases and related initiatives to aid data preservation, sharing and re-use.
- An international list of data repositories is available via Databib or Re3data.

5- Archiving and preservation (including storage and 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 approximated end volume, what the associated costs are and how these are planned to be covered.

Questions to consider on Preservation Plan:

- 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 how datasets that have 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 you 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 will 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.

Annex 3: GenTree Data Sharing Agreement

1. Conditions for supplying Data and Metadata to the GenTree Information System

1.1 Data and Metadata coming from Parties of the GenTree Project

a) Data and Metadata defined as Background

- Each Party shall supply the GenTree Information System with its Data and Metadata as soon as possible, if these Data and Metadata are needed for carrying out the activities of the research project, unless these Data or Metadata have been excluded or limited in the Consortium Agreement.
- The Parties will provide a citation (acknowledgement) reference for each Data set included as a Metadata into the GenTree Information System

b) Data and Metadata (Foreground) produced by the Parties

- Each Party shall supply the GenTree Information System with its Data and Metadata as soon as they are generated if these Data and Metadata are needed for carrying out the activities of the research project.
- The Parties will provide a citation (acknowledgement) reference of each Data set provided into the GenTree Information System

1.2 Data and Metadata coming from third Parties

- The Parties shall supply the GenTree Information System with Data and Metadata from third parties, according to the rules defined by each such third party.

2 Access policy to Data and Metadata during, outside and after the GenTree Project

2.1 Access to Metadata

- Public access to the Metadata will be given once these are entered into the GenTree Information System.

2.2 Access to Data

a) Access to Data defined as Background

- Access to Data defined as Background will be given to the GenTree partners as needed for the implementation of their research activities and projects in compliance with the Consortium Agreement.
- Access to Data defined as Background will be given to the scientific community upon prior authorisation given by the Party who owns the data or rights pertaining to such data

b) Access to Data defined as Foreground

- Access to Data defined as Foreground obtained within the GenTree Project, will be given to the GenTree partners once they are entered into the GenTree Information System, and to the public, after a maximum delay of 18 (eighteen) months after the end of GenTree.

- Parties who own the Data will endeavour to shorten this period of 18 (eighteen) months. The period of 18 (eighteen) months can be extended through a decision of the GenTree ExCom, after prior consultation of the owner the Data.

3 Publication/communication

3.1 Metadata public release

- Metadata shall be released to the Public as soon as possible. Any publication or communication of Data and/or Metadata shall be made according to the provisions of article 8.4 (Dissemination) of the Consortium Agreement.

3.2 Referencing and acknowledgement

- Any publication of the results obtained within the GenTree project must reference the project as follows: “This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 676876”.
- Any publication of results obtained that use Data and/or Metadata shared within the GenTree Information System, must reference the Data and/or Metadata owner, the citation reference of the dataset (whenever provided in the metadata) and the project.
- Any publication of a synthesis of Data and/or Metadata shared in the GenTree information system, including for dissemination or training, must reference the Data and/or Metadata owner, the citation reference of the datasets (whenever provided in the metadata) and the project.

Annex 4: Creative Commons Licenses

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 neighboring 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.	View description
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.	View License Deed View Legal Code
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.	View License Deed View Legal Code
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.	View License Deed View Legal Code
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.	View License Deed View Legal Code
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.	View License Deed View Legal Code
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.	View License Deed View Legal Code