# DMP du projet "Grape Genes for WAter Scarcity"

Plan de gestion de données créé à l'aide de DMP OPIDOR, basé sur le modèle "INRA - Trame générique projet" fourni par INRAE - Institut national de recherche pour l'agriculture l'alimentation et l'environnement.

# **Renseignements sur le plan**

Titre du plan	DMP du projet "Grape Genes for WAter Scarcity"	
Version	Version intermédiaire	
Domaines de recherche (selon classification de l'OCDE)	Agriculture, forestry, and fisheries	
Langue	fra	
Date de création	2019-02-18	
Date de dernière modification	2022-05-20	
Identifiant	DMP-ANR G2WAS	

# Renseignements sur le projet

Titre du projet	Grape Genes for WAter Scarcity
Acronyme	G2WAS
Acronyme Résumé Sources de financement	In perennial species, yield and production quality are impacted by water stress with marked interannual effects. The knowledge of the physiological and genetic mechanisms regulating grapevine responses to WD remain largely insufficient to adapt the viticulture to climatic challenges. Most often, responses to water stress have been studied during a single vegetative cycle, considering traits independently and using a limited range of genetic diversity. The G2WAS project aims to study the responses of grapevine to water deficit on intra- and inter-annual scales, by integrating the dynamics of production, storage and utilization of carbon resources in both vegetative and reproductive systems. This study will be performed with a diversity panel designed to maximize the genetic diversity of the cultivated species (V. vinifera). In order to decipher the genetic and physiological bases of adaptation to drought, and to incorporate them into breeding programs, several innovative approaches will be run: i) advanced phenotyping of vegetative and reproductive organs targeted at several critical developmental stages, with a focus on carbon allocation, ii) identification by exhaustive transcriptomics (RNAseq) of co-regulated gene networks; iii) genotype-phenotype whole genome association (GWAS) analysis applied to a panel of 279 varieties iv) development of a multi-trait and multi-year statistical model to improve prediction accuracy. Performed for the first time in perennials, such a combination of methods will improve the detection of QTL and the prediction of individual genetics rules. This multidisciplinary approach will be supported by the G2WAS consortium which brings together specialists in eco-physiology, physiology, quantitative and functional genetics, statistics and breeders. In addition to the coordination (WP1), the project is based on 4 WP: a physiological study (WP2) of 16 contrasting genotypes confronted to a gradient of 10 hydric conditions under tightly controlled environment (PhenoDyn platform) to provide
Sources de financement	• ANR : ANR GENERIQUE 2019
Date de début	2019-10-01
Date de fin	2024-07-31
Partenaires	<ul> <li>Institut National de la Recherche Agronomique <u>grid.414548.8</u></li> <li>Institut National d'Enseignement Supérieure en Agriculture, Alimentation et Environnement</li> <li>AgroParisTech</li> </ul>

Produits de recherche :

1. Default research output (Jeu de données)

# Contributeurs

Nom	Affiliation	Rôles
Cédric Goby		• Responsable du plan de gestion de données
Laurent Torregrosa - <u>https://orcid.org/0000-0003-1622-5610</u>		<ul> <li>Coordinateur du projet</li> <li>Personne contact pour les données</li> </ul>

Droits d'auteur :

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous

n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

## Informations sur le plan de gestion

Auteur du plan de gestion s'il est différent du Principal Investigator/Researcher

Cédric Goby and Vincent Nègre

### Affiliation de l'auteur du PGD

Institut National de la Recherche Agronomique, France

## Date de création du PGD

01/02/2019

#### Version en cours

2.0

## Date de la dernière version

10/04/2019

# Informations sur le projet

## Identifiant de l'appel à projet (call for proposal)

ANR AAPG2019 - G2WAS - Grape Genes for WAter Scarcity

## Financeur(s) du projet

Pour les projets européens on pourra se reporter aux informations telles qu'elles apparaissent dans <u>Cordis</u>. Le <u>répertoire CrossRef des financeurs</u> peut être utilisé pour indiquer le libellé et l'identifiant DOI attribués à un financeur par CrossRef (exemple European Commission http://dx.doi.org/10.13039/501100000780 pour la Commission Européenne).

Agence nationale de la recherche (ANR) - French Research Agency

## Nom du programme de recherche

Pour les projets européens on pourra se reporter aux informations telles qu'elles apparaissent dans Cordis

Acronym: G2WAS - Grape Genes for WAter Scarcity

Long title: Genetic architecture of the tolerance to water deficit in a perennial fruit species (V. vinifera)

### Référence de la convention de financement

AAPG2019

## Acronyme du projet

G2WAS

### Nom du projet de recherche

Si le projet correspond à un financement lié à un appel à projet, nom du projet tel qu'il apparait dans la réponse à l'appel à projet.

G2WAS - Grape Genes for WAter Scarcity

Institution leader du projet, coordinateur bénéficiaire (nom, pays)

Nom de l'institution tel qu'il est connu par le financeur. On pourra se reporter à la base de données <u>GRID Global Research Identifier Database</u> ou pour la France au <u>RNSR Répertoire</u> National des Structures de Recherche Institut National de la Recherche Agronomique, France

Autres partenaires (nom, pays, rôle de chacun des partenaires en dehors de l'institution leader)

Montpellier SupAgro - Institut National d'Etudes Supérieures Agronomiques de Montpellier AgroParisTech IFV - Institut Français de la Vigne et du Vin

Unité de rattachement du responsable du projet

Suivre la note de service Inra 2016-13 sur la signature "monoligne" des publications ou la charte des publications de l'Alliance nationale pour les sciences de la vie et de la santé

Montpellier SupAgro

Dates et durée du projet

58 months (1/10/219 to 31/7/2024)

## Présentation succincte des données du projet

Brève présentation des données générées, collectées ou réutilisées :

- périmètre, type, échelle
- origine : collecte de nouvelles données ou réutilisation de données (conversion / transformation de données ; partage / échange de données ; achat de données), publications associées

Identification of Partners in charge of data production : P1: INRA and SUPAGRO (AGAP) P2: INRA and SUPAGRO (LEPSE) P3: IFV

### Data description by category

#### 1. Genotypic data 1.1 GBS (reused data\*): P1

63 kSNPs for 279 grapevine varieties, from Flutre et al (in preparation and ICGBG Bordeaux 2018), to be used for GWAS in WP4, in combination with new genotypic and phenotypic data (see below) \*will be open when published

### 1.2. Capture (new data): P1

20-40 kSNPs for 279 grapevine varieties, to be generated by capture in regions insufficiently covered, with QTLs related to response to drought, or with genes differentially transcribed under drought, to be produced in WP3 and used in WP4 for GWAS, in combination with already available genotypic data (see above) and new phenotypic data (see below) 2. Transcriptional data: P1

2.1. Microvine berry development (reused data)

RNAseg done for 3 microvines lines and 7 developmental stages of berry development.

These data will be free for research as soon published and under Poupelain/SupAgro agreement. 2.2 Transcriptional data (new data): P1 and P2

RNAseq expression data for 6 grapevine genotypes showing specific physiological responses to WD, under 2 WD conditions, on 2 organs (fruits and shoots), with 3 replicates, to be produced in WP2 3. Phenotypic data (new data): P1 and P2

For 16 grapevine genotypes under 10 watering conditions, to be produced in WP2: P1 and P2  $\,$ 

- Vegetative traits: interannual C allocation (carbohydrates content in different organs) starting at véraison (before stress application), daily transpiration, midday gas exchange, chlorophyll fluorescence, predawn- and midday- leaf water potential, shoot growth between the beginning and the end of the Phenodyn experiment (by image analysis)

- Reproductive traits: yield components at maturity (cluster number and weight, berry weight), cluster growth at least weekly, berry development and composition at two stages (sugars, acids, ions, anthocyanins and proactions, fruit respiration, berry temperature For 279 grapevine varieties under 2 watering conditions, with 2 replicates, to be produced in WP3: **P1 and P2** 

- Vegetative traits: interannual C allocation (carbohydrates content in different organs) starting at véraison (before stress application), daily transpiration, shoot growth weekly (by image analysis), photosynthetic activity twice a month (N content with SPAD, specific leaf area on leaf discs)

- Reproductive traits: yield components at maturity (cluster number and weight, berry weight), cluster growth weekly, berry development and composition at two stages (sugars, acids and ions contents) 4. Seeds and information about 5 varieties in selection (WP5) : P1, P2 and P3

Previous data + new genotypic data : P3 Previous agronomic data : P3 New genotypic and phenotypic data : P1, P2 and P3

# Droits de propriété intellectuelle

Qui détiendra les droits sur les données et les autres informations créées lors du projet ?

Faire attention guand un partenaire privé amène des données dans le projet.

A consortium agreement has been signed to define the intellectual property rights, exploitation rights, on the results of the project. General principles:

Previous data remain the property of the providing partner (P1, P2, P3).

Data generated and results obtained by only one partner belong to the partner that generated them.

Data generated and results obtained by several partners are equally owned by the partners who generated them.

Du matériel protégé par des droits spécifiques sera t-il utilisé au cours du projet ? Dans ce cas, qui s'occupe des formalités à accomplir, obtient les autorisations d'utilisation et de diffusio

Plant material from IFV (5 varieties in WP2 and WP5) and INRA (panel of 279 varieties from the Vassal-Montpellier Grapevine Biological Resources Center in WP3) will be used in the project. A consortium agreement will specify the rights to use the plant material, and the formalities to be carried out when required.

## Confidentialité

Identification des jeux de données confidentielles

Genotypic and phenotypic data sets on the 5 genotypes of the private partner IFV will be considered confidential. A consortium agreement has been signed.

Quelles sont les mesures prises et les normes auxquelles il est nécessaire de se conformer pour garantir cette confidentialité ? Les règles de confidentialité doivent être écrites et diffusées auprès des utilisateurs. Une charte peut éventuellement être imposée engageant l'utilisateur à respecter ces règles.

A confidentiality clause will be included in the consortium agreement for this material.

Le cas échéant, comment la confidentialité de données fournies par des personnes sera garantie lorsque les données seront partagées ou rendues disponibles pour une analyse de second niveau ?

Voir NGUYEN, B. (2015). Techniques d'anonymisation. Statistique et Société, 2(4)

Data will be secured on a dedicated server, with password restricting the access and communication encryption. Access to these data will be examined on request following rules defined in the consortium agreement.

# Partage des données à l'issue du projet

Y a t'il une obligation de partage (ou à l'inverse une interdiction ou une restriction)

No obligation of public access as one of the partners is of private status. Nevertheless, as the consortium is mainly formed with public institutions, it will be willing to widely share as much as possible the data. A consortium agreement has been signed.

Quelles données seront partagées à l'issue du projet ? Si toutes les données ne sont pas disponibles de la même façon, ou en même temps, le préciser

For potential commercial exploitation by the Partner IFV, genotyping and phenotyping data for the 5 varieties of the IFV will not be shared at the end of the project unless otherwise agreed by the IFV. The other data will be shared following rules defined in the consortium agreement.

Quelles sont les réutilisations potentielles de ces données ?

Reuse of metadata on collected samples for further sample analyses in other projects. Comparison of the data obtained in the G2WAS project with data obtained in subsequent projects using the same panel (e. g. Muse Panel 279 2020-2030).

La lecture des données nécessite-t-elle le recours à un logiciel ou un outil spécifique ? Si oui, lequel et comment y accéder ?

No.

Comment les données seront-elles partagées ?

Comment les données sont-elles rendues accessibles, par exemple : dépôt dans un entrepôt (le choix d'entrepôt est abordé dans la section "Archivage et conservation des données à long terme"). Le portail Data Inra peut accueillir les jeux de données Inra.

The data will be accessible from the INRA data repository : <u>https://data.inra.fr/</u>

Avec qui ? sous quelle licence ?

Procédure d'accès : ouverture à tous ou à un groupe spécifique, type de contrôle. Voir : <u>Choisir une licence</u>

The choice of the most appropriate license will be finalized in the consortium agreement within 3 months after the start of the project (see GANTT).

A partir de quand ?

The choice of the most appropriate license will be finalized in the consortium agreement within 3 months after the start of the project (see GANTT).

Pendant combien de temps ?

S'assurer que les métadonnées seront accessibles même si les données ne le sont plus.(par exemple en déposant dans un entrepôt de confiance, en identifiant les données par un DOI...)

The INRA data repository (https://data.inra.fr/) which will be used in this project will ensure a long term access to metadata.

Les données seront-elles identifiées par un identifiant pérenne (DOI ou autre) ?

A DOI is automatically assigned to the datasets deposited in the INRA data repository.

Quel est l'organisme qui se chargera de la demande d'identifiant dans le cas de projets multi-partenaires ?

Si l'Inra fait se charge de la demande de DOI, vous pouvez utiliser le service d'attribution de DOI de l'Inra. A noter qu'un DOI est automatiquement attribué aux données déposées dans l'entrepôt <u>Data Inra</u>

INRA will be in charge of the DOI's requests

## Description et organisation des données

Quels méthodes et outils sont utilisés pour acquérir et traiter les données ? Précisez les différents formats dans lesquels les données seront disponibles aux différentes phases de la

Data will be stored in the information system PHIS (Neveu et al. New Phytologist 2018).

PHIS is an open source Information System for phenomic data that allows integration and sharing of multi-source and multi-scale data as well as semantic annotation of experiments with knowledge and metadata.

PHIS can dialogue and interoperate with external resources.

For instance, PHIS currently provides elaborated data to the Environment and Phenotype Information System (EPHESIS, https://urgi.versailles.inra.fr)/ephesis/), to the collaborative Breeding API (BrAPI) (http://braj The software MySQL is also used to build the database.

Documentation associée aux données

A "README" file will be written to gather basic information on the data and metadata (name of the source, file format, identifier, description of the content, environmental conditions when obtaining the mples, type of analytical material used...)

Quels types de métadonnées seront produites pour accompagner les données ? Quels sont les standards et les vocabulaires ou taxonomies qui seront utilisés pour décrire les données ?

Metadata for WP2, 3 and 5 include protocols, description of variables, plant material and environmental conditions, time of day as well as electronic notebook. Metadata will be stored in the PHIS information system. PHIS uses ontologies publicly available through portal such as agroportal (http://agroportal.lirmm.fr/) or crop ontology portal (http://www.cropontology.org/). The software MySQL is also used to build the database

Comment les métadonnées seront elles produites ?

Il existe des outils pour produire ces métadonnées

PHIS hosts both metadata, raw data and processed data. The software MySOL is used to build the database.

Comment les fichiers de données sont-ils gérés et organisés au cours du projet : contrôle des versions, conventions de nommage des fichiers, organisation des fichiers...

Séparer les données brutes des données traitées, des délivrables et des programmes de traitement. Définir des règles d'organisation et de nommage des fichiers de données. En fournir éventuellement une représentation graphique Ne pas modifier les données brutes : toute modification doit entraîner un changement de nom des fichiers

Voir <u>Nommage et organisation des fichiers de données</u> (site Gestion et partage des données scientifiques).

Part of raw data (online) will directly integrate PHIS. Other raw data (offline) will be first deposited in a collaboration platform (Nextcloud) hosted on INRA dedicated servers. Owncloud ensures file versioning. Processed data will be stored separately. Data processing workflows will also be stored. Files will be organised following workpackages and tasks as described in the project.

### Quelle est la procédure de contrôle qualité des données ? joindre éventuellement le plan d'assurance qualité

Biological material will be labelled using QRcodes corresponding to a unique ressource identifier (URI). URI are recorded in the information system. Events (management, faults, meteo, ....) could also be integrated into the information system to keep tracability on data production

## Stockage et sécurité des données

Stockage : Quels seront les supports utilisés pour les données au cours du projet ?

Due to the diversity of traits and phenotyping tools, offline raw data will be primarily stored on different devices (laptop, desktop, tablet ; USB key ; smartphone ; sensors; loggers). Generated files will be regularly deposited on the collaboration platform (Nextcloud). Data will be stored on a cloud S3 Object Storage. To prevent data loss and allow continuity in operations and services the S3 Object Storage will be replicated in a second datacenter. A backup system will be installed on a Linux backup server in a second datacenter to make incremental backup of the data and allow recovery procedure.

Stockage : Quels seront les types de flux empruntés par les données au cours du projet ?

Fournir éventuellement un schéma fonctionnel du système d'information

Online data in phenotyping platforms will automatically integrate PHIS using web services. Offline data will be first deposited by each partner in a collaboration platform (Nextcloud) hosted on INRA dedicated servers.

Users will be provided a password or a private key to connect to the collaboration platform. Only secure protocols are used to upload data to the application server : https, sftp, ssh

Stockage : Quelle est la volumétrie prévisionnelle ? En mesure de stockage. Elle pourra être réévaluée au cours du projet.

2.5 Teraoctets for RNAseq data

7.5 Teraoctets for phenotypic data (including pictures from semi-automatic phenotyping)

Stockage : Où sont hébergées physiquement les données, sur quel type d'hébergement ?

The data of the project will be replicated in two different datacenters managed by INRA IT Services. Data are stored on a cloud S3 Object Storage.

Stockage : Où sont localisées géographiquement les données ?

Data are replicated in two different INRA data centers in France : Jouy-en-Josas and Toulouse.

Sécurité : L'entité hébergeant physiquement les données a-t-elle une politique de sécurité pour son système d'information ? La Politique Sécurité des Systèmes d'Information (PSSI) est un ensemble de règles de sécurité faisant référence pour l'entité considérée

Each datacenter has a high level of security : building access cards, CCTV, hardware and electrical equipment redundancy, network monitoring, 24/7 support.

Sécurité - Confidentialité : les données feront-elles l'objet d'échange ou de partage avec de tiers acteurs ?

Confidentialité : propriété selon laquelle l'information n'est pas rendue disponible ni divulguée à des personnes, des entités ou des processus non autorisés (ISO 27000). Voir aussi "Renseignements confidentiels" sur le glossaire Research Data Canada (RDC).

During the project, no sharing with third parties, except with the agreement of the partners and according to the consortium agreement.

Sécurité - Confidentialité : comment sont déterminés les droits d'accès aux données pendant les recherches ?

Gestion des accès : Il convient que les propriétaires des actifs déterminent des règles de contrôle d'accès, des droits d'accès et des restrictions d'accès appropriés aux fonctions spécifiques de l'utilisateur des actifs. Les contrôles d'accès sont à la fois logiques et physiques (<u>ISO 27002</u>).

The project manager will maintain a list of users and groups. He will define the access level for data for each user and groups.

Sécurité - Confidentialité : De quelle manière l'ensemble des chercheurs partenaires du projet auront-ils accès aux données pendant la recherche ? Méthode d'identification, d'authentification.

Access to data will be granted by using two authentication methods : password or public-private keys.

Sécurité - Intégrité - Tracabilité : Quelles sont les mesures de protection mises en œuvre pour suivre la production et l'analyse des données au cours du projet ?

Intégrité : propriété d'exactitude et de complétude (ISO 27000). Les données doivent être celles que l'on attend, et ne doivent pas être altérées de façon fortuite, illicite ou malveillante. En clair, les éléments considérés doivent être exacts et complets. (Wikipédia juin 2016)

QRcodes for biological material and identification of sensors and operators will ensure tracability of raw data production. Offline data transfer to storage tools will be tracked through the password or the public-private key which will be required by the collaboration tool. File versioning will make possible the control of irregular deposition. Mechanisms will protect data against fraudulent access (data transfer encryption). Regular backup will prevent data loss.

# Archivage et conservation des données après la fin du projet

Quelles sont les données à conserver sur le moyen ou le long terme et quelles sont les données à détruire ?

A large storage space has been anticipated to store project data, including data acquired on semi-automatic phenotyping platforms (pictures and scans) and RNA resequencing data. The compression of automatically acquired images may be considered after discussion with partners.

Sur quelle plateforme d'archivage pérenne seront archivées les données à conserver sur le long terme ? Sinon, quelles procédures seront mises en place pour la conservation à long terme ?

Il s'agit ici de plateformes d'archivage pérennes destinées à pérenniser les données, comme le C.I.N.E.S. Les entrepôts de données ne possèdent, à quelques exceptions près, pas cette possibilité.

Long-term storage will be handled by C.I.N.E.S (Centre Informatique National de l'Enseignement Supérieur)

Quelle est la durée de conservation des données ?

20 years

Qui sera responsable de la conservation à long terme ? nommer un contact individuel.

Le coordinateur du projet est responsable de la gestion des données durant le projet et de leur archivage à la fin de celui-ci. Le directeur de l'unité ayant coordonné le projet peut être responsable sur le long terme.

Laurent Torregrosa

Quel sera le volume de ces données ?

10 Teraoctet in total (estimation)

Quelles garanties de financements couvriront les coûts associés à la conservation à long terme ?

UK Data Service - Data management costing tool and checklist apporte des éléments pour estimer le coût de la gestion des données

Data will be stored on INRA public servers with funding from the budgets of the units concerned.