

➤ Retour d'expérience : Mise en place d'un Plan de Gestion de Données (PGD) de structure à l'échelle d'une UMR (UMR 1131 SVQV)

Amandine Velt – Référent données opérationnel (RDO) de l'UMR 1131 SVQV – INRAE Grand Est - Colmar

UMR 1131 SVQV (Santé de la Vigne et Qualité du Vin)

L'UMR 1131 Santé de la Vigne et Qualité du Vin est une unité mixte de recherche INRAE - Université de Strasbourg, située à Colmar.



Objectif : développer des stratégies durables pour protéger la vigne de ses principaux bioagresseurs, tout en préservant la qualité des raisins et des vins.

Trois équipes de recherche:

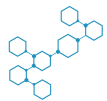
- **Génétique & Amélioration de la Vigne** (GAV) : résistance aux maladies, création variétale, phénotypage
- **Virologie & VEction** (ViVe): interactions plante-virus, vecteurs, imagerie
- **Génomique & Métabolisme de la Vigne** (GMV) : génomique, composés aromatiques

Types de données manipulées au sein de l'UMR SVQV



Données de génomique et de séquençage

- Séquençage haut-débit (FASTQ)



Données de métabolomique

- Spectrométrie de masse (formats Thermo/Agilent)



Données d'imagerie

- Images « Phénomobile » (LIDAR, caméras, GPS - fichiers HDF5)
- Images de microscopie, capteurs NIRS (CZI, TIFF, JPEG, TXT)



Données de phénotypage

- Plateforme de phénotypage - VEGOIA (TXT)
- En dehors de la plateforme VEGOIA (TXT)



Données spécialisées

- entomologie (JPEG, PDF, XLS)
- SSR, GBS (FSA, FASTQ)

Qu'est ce qu'un RDO à INRAE

Le réseau des **R**éférents **D**onnées **O**opérationnels :



250 agents issus des **14** départements scientifiques d'INRAE et répartis sur environ **270** unités



Mission : assister les collectifs scientifiques INRAE dans la gestion des données de la recherche



Lettre de mission de 5% de mon temps de travail



INRAE

Plan de Gestion de Données de structure de l'UMR 1131 SVQV

30/04/2025 / Amandine Velt

Plan de Gestion de Données (PGD)

Document établi en début de projet décrivant comment les données seront collectées, documentées, utilisées, stockées et partagées.



Produire des données **FAIR** : Faciles à trouver, Accessibles, Interopérables et Réutilisables



<https://dmp.opidor.fr/>

INRAE

Plan de Gestion de Données de structure de l'UMR 1131 SVQV

30/04/2025 / Amandine Velt

Deux types de modèles de PGD proposés par INRAE : « projet » et « structure »



Projet

Modèle de plan de gestion de données adapté aux projets de recherche.



Structure

Modèle adapté à la gestion des données dans tout type de structure (unité, plateforme,...).

Avantages du PGD de « structure »



Fixer la politique de données
de la plateforme ou de l'unité



Gagner du temps,
document de référence



Faciliter la rédaction
de PGD de projet



Élément de qualité des données

PGD de structure pour l'UMR 1131 - SVQV

L'UMR SVQV génère, stocke des données très hétérogènes.



Un PGD permet de :

- identifier clairement toutes les données générées par l'UMR,
- décrire les procédures de gestion des données au sein de l'UMR,
- répondre aux exigences des agences de financement.



Mode opératoire pour la rédaction du PGD de structure



1. Réunion initiale

- Rencontre avec les responsables des équipes GAV/GMV/VIVE
- Inventaire des données générées
- Identification des personnes contacts



2. Entretiens individuels

- Rendez-vous avec chaque personne « contact »
- Rédaction collaborative du PGD



3. Relecture collective

- Envoi du PGD à toutes les personnes « contact » et aux responsables d'équipe
- Correction et ajustements



4. Validation finale

- Présentation du PGD en assemblée générale de l'UMR
- Validation officielle du document

→ Pour en savoir plus sur la rédaction de PGD :

INRAE

Plan de Gestion de Données de structure de l'UMR 1131 SVQV

30/04/2025 / Amandine Velt



Constats et pistes d'amélioration pour la gestion de données

La rédaction du PGD a permis d'identifier des besoins communs et des écarts entre équipes et plateformes.

Principaux constats :

-  Données dispersées : parfois stockées localement sur les ordinateurs des agents ou des instruments.
-  Métadonnées absentes ou incomplètes : difficile de partager, comprendre ou réutiliser les données.
-  Organisation hétérogène : chaque agent applique sa propre logique de classement.
-  Pas d'archivage à long terme : risque de perte d'information.
-  Bonnes pratiques non partagées : (contrôle qualité, fichiers Excel de suivi) mais pas documentées ou dépendantes d'une seule personne.
-  Pas de vision globale : difficulté à mutualiser les données produites.

Les informations écrites pour chaque type de données

1

Informations sur la structure

2

Présentation générale des données

3

Droits de propriété intellectuelle

4

Sensibilité des données

5

Partage des données

6

Organisation et documentation des données

7

Stockage et sécurité des données

8

Archivage et conservation des données

Stockage sécurisé des données « légères » de l'UMR 1131 SVQV



Home utilisateur INRAE

Stockage personnel sécurisé sur l'ordinateur de chaque agent (datacenter de Paris).

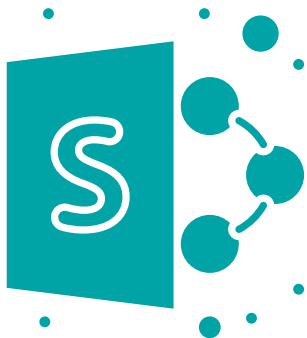
→ Usage : données légères (Excel, PDF, Word, etc.).



Nextcloud INRAE

100 Go d'espace personnel par agent, synchronisable avec son poste local.

→ Idéal pour les sauvegardes quotidiennes.



SharePoint INRAE

Espace commun pour l'UMR, type site web.

Stockage des données « lourdes » de l'UMR 1131

SVQV – solution locale

Bâtiment principal
Serveur “calcul”



Bâtiment « Vigne et Vin »
Serveur “backup”



Une synchronisation
hebdomadaire

● Cluster de 6 nœuds

● 162 CPUs

● 1,5 To RAM

● 83 To de stockage

● 103 To de stockage

5767 euros -
extension garantie 5 ans

7555 euros -
extension garantie 7 ans

INRAE

Plan de Gestion de Données de structure de l'UMR 1131 SVQV

30/04/2025 / Amandine Velt

Stockage local : une solution moins chère mais fragile



Besoin identifié

- Solution « datacenter » professionnelle pour garantir maintenance, stockage et pérennité.



Frein principal

- Coût élevé pour une petite UMR (offres datacenters nationaux payantes).



Ressources limitées chez nous

- Seulement 2 agents pour la gestion des serveurs en plus de leur activité de recherche.
- Une salle serveur peu adaptée, équipée d'une armoire de serveurs et d'une climatisation dédiée au refroidissement de la pièce.



Coûts Unistra/DNUM

- Stockage S3 (sans sauvegarde) : 45 €/To/an
- Commvault (avec sauvegarde) : 36 €/To/an
- → Pour 100 To sauvegardés : 8100 €/an, contre 13322 € pour l'achat de 2 NAS garantis entre 5 et 7 ans.



INRAE

Plan de Gestion de Données de structure de l'UMR 1131 SVQV

30/04/2025 / Amandine Velt

Entrepôts utilisés pour le partage des données de recherche liées à une publication scientifique



Données de séquençage — ENA

- Dépôt obligatoire pour les publications en génomique.
- Standard international.
- Données accessibles mondialement, mais stockées sur serveurs européens.
- Conformité : ouverture, traçabilité (DOI attribué).
- Métadonnées strictes, standardisées et vérifiées.



Autres données — Entrepôt Data INRAE

- Dépôt via data.gouv.fr - Data INRAE
- Conformité : ouverture, traçabilité (DOI attribué).
- Métadonnées libres, contrôle limité.

Archivage des données de recherche : quelles solutions pour l'UMR 1131 SVQV ?

Constat

- Dépôt actuel dans des entrepôts publics (ENA, Data INRAE).
- Pas de solution d'archivage électronique pérenne.

Enjeux

- Conservation très long terme des données.
- Garantie d'intégrité, traçabilité et accessibilité.
- Respect des obligations légales et institutionnelles.

Pistes d'action

- Collaborer avec le Service des archives de l'Université de Strasbourg.
- Évaluer des solutions d'archivage électronique conformes aux normes.
- Intégrer l'archivage dans le Plan de Gestion des Données (PGD) de l'UMR.

Conclusion et perspectives

Bilan

- PGD disponible en interne (format PDF).
- Rédaction : 1 an (2 jours max/mois).
- Publication prévue sur Data INRAE.

Avantages

- Conformité renforcée aux exigences.
- Meilleure valorisation des données.
- Document de référence pour l'UMR.
- Aide pour la rédaction des futurs PGD projets.

Perspectives

- Sensibilisation/formation à la gestion des données.
- Mise à jour annuelle du PGD.
- Réflexion sur l'archivage pérenne.

Merci pour votre attention

