



**HAL**  
open science

## Plan de gestion de données de l'Unité Expérimentale INRAE du Pin

Frédéric Launay, M. Gaborit, Annabelle Bachelet, Frédéric Herault

► **To cite this version:**

Frédéric Launay, M. Gaborit, Annabelle Bachelet, Frédéric Herault. Plan de gestion de données de l'Unité Expérimentale INRAE du Pin. INRAE. 2024. hal-04728765

**HAL Id: hal-04728765**

**<https://hal.science/hal-04728765v1>**

Submitted on 9 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle « INRAE – Modèle Structure » fourni par INRAE – Institut national de recherche pour l'agriculture, l'alimentation et l'environnement

## Plan de Gestion de données Structure

UEP : Unité Expérimentale du Pin



Version 4  
30/08/2024

## Table des matières

I-	Renseignements sur la structure.....	2
	Description de la structure.....	2
	Partenaires de la structure.....	2
	Responsabilité dans la structure.....	3
II-	Renseignements sur le plan.....	3
	Objectifs du Plan de Gestion de Données (PGD).....	5
	Produits de recherche décrits dans ce PGD.....	5
III-	Présentation Générale des données.....	5
	1) Ensemble des données issues de mesures ou observations effectuées sur les animaux.....	5
	Mode d'obtention des données.....	5
	Origine.....	6
	Type de données.....	7
	Nature des données.....	8
	Format des données.....	8
	Périmètre thématique des données.....	9
	2) Ensemble des données issues de mesures ou de relevés de l'environnement d'élevage.....	9
	Mode d'obtention des données.....	9
	Origine.....	9
	Type de données.....	10
	Nature des données.....	10
	Format des données.....	11
	Périmètre thématique des données.....	11
IV-	Droits de propriété intellectuelle.....	11
V-	Sensibilité des données.....	11
VI-	Partage des données.....	12
VII-	Organisation et documentation des données.....	13
	Stockage et sécurité des données.....	16
VIII-	Archivage et conservation des données.....	17

## I- Renseignements sur la structure

UEP 0326 – Unité Expérimentale du Pin

- Unité Expérimentale
- ISC (Infrastructure Scientifique Collective)
- Infrastructure de recherche (IR)
  - L'unité expérimentale est intégrée à deux IR : LIPH4SAS (Infrastructure nationale dédiée au phénotypage des animaux d'élevage : <https://liph4sas.fr/> ) et AgroServ (Infrastructure du programme Horizon Europe (2022 – 2027) accompagnant la transition agroécologique de l'agriculture en Europe : <https://agroserv.eu/agroserv-homepage> )

**Identifiant de la structure** : identifiant RSNR 196217742E

**Etablissement(s) tutelle(s)** : INRAE

**Département de rattachement INRAE (ou anciens départements Inra)** : GA = Génétique animale

**Financier(s) (permettant l'acquisition des jeux de données – hors projet)** :

- INRAE
- Conseil Régional de Normandie
- Union Européenne : Fonds européen de développement régional (FEDER)
- Agrial : coopérative agricole et agroalimentaire
- SYNERGI : entreprise de conseil en élevage et fourniture de matériel d'élevage

### Description de la structure

L'Unité INRAE du Pin est une ferme expérimentale laitière et fourragère, à mission nationale et européenne. Des travaux de recherches fondamentales et appliquées y sont menés tout au long de l'année, en collaboration avec des équipes scientifiques de INRAE, de l'enseignement supérieur et des instituts techniques et des entreprises privées ou publiques.

L'unité est rattachée administrativement au Centre INRAE Bretagne Normandie et scientifiquement au département de Génétique Animale (GA).

Au cœur du bocage normand, dans l'Orne, l'unité s'organise autour d'un pôle laitier réparti sur 2 sites distants de 2 kilomètres (avec un troupeau de 600 têtes, dont 250 vaches à la traite) sur un total de 340 hectares dont 300 hectares en herbe.

Pour plus d'informations : <https://annuaire.inrae.fr> -Unité Expérimentale du Pin

Site internet : <https://uep.isc.inrae.fr/qui-sommes-nous/l-unite-experimentale-du-pin>

### Partenaires de la structure

- Le Département de Génétique Animale (GA) – Identifiant RSNR : 200319711S
- Le Département de Physiologie Animale et Système d'Élevage (PHASE) - Identifiant RSNR : 200417731K
- Le Département de Santé Animale (SA) - Identifiant RSNR : 199817728H
- Unité Mixte de Recherche sur les Herbivores - Identifiant RSNR : 199917863A

- Unité de Génétique Animale et Biologie Intégrative - Identifiant RSNR :200917790C
- Unité de Physiologie, Environnement et Génétique pour l'Animal et les Systèmes d'Élevage - Identifiant RSNR :201220380<sup>E</sup>

## Responsabilité dans la structure

Nom Prénom	Courriel	Rôle
LAUNAY Frédéric	frederic.launay@inrae.fr	Directeur Unité
GABORIT Marie	marie.gaborit@inrae.fr	Cheffe d'Élevage - Directrice Adjointe
BACHELET Annabelle	annabelle.bachelet@inrae.fr	Responsable expérimentations

## II- Renseignements sur le plan

Titre du plan	"Plan de Gestion des données Entité de Recherche de l'unité expérimentale du Pin"
Livrable	PGD de structure V1
Objet/périmètre du plan	<p>Le plan de gestion de données de l'unité est un document cadre décrivant le cycle de vie des données produites ou utilisées dans le collectif.</p> <p>Il se décline en 2 produits de recherche regroupant l'ensemble des typologies de données produites ou mobilisées au sein de l'unité. Ces deux produits de recherches concernent :</p> <ul style="list-style-type: none"> <li>- Les données issues de mesures ou observations effectuées sur les animaux</li> <li>- Les données issues de mesures ou de relevés de l'environnement d'élevage</li> </ul> <p>Il pourra servir de référence pour l'élaboration de plan de gestion de données de projet de recherche s'appuyant sur les travaux menés dans l'unité expérimentale.</p>
Domaines de recherche (selon classification de l'OCDE)	Sciences animales et élevage laitier, agriculture, silviculture et pêche
Langue	français
Date de création	2022-11-30
Date de dernière modification	2024-07-11
Licence	Creative Commons Attribution Non-Commercial Share Alike 4.0 International (CC BY-NC-SA 4.0)
Documents (publications, rapports, brevets, plan expérimental...), sites web associés	<p><u>Données météorologiques</u> : <a href="https://agroclim.inrae.fr/climatik/">https://agroclim.inrae.fr/climatik/</a></p> <p><u>Site internet</u> : <a href="https://uep.isc.inrae.fr/qui-sommes-nous/l-unite-experimentale-du-pin">https://uep.isc.inrae.fr/qui-sommes-nous/l-unite-experimentale-du-pin</a></p>

<i>Plans de gestion en lien avec le projet</i>	<p><u>Plan de gestion de données du CATI</u>  SICPA : <a href="https://doi.org/10.57745/S19P9X">https://doi.org/10.57745/S19P9X</a></p> <p>Plan de gestion de données de l'unité Agroclim (en cours de validation)</p> <p>Plan de gestion de données du CTIG (Centre de Traitement de l'Information Génétique) (en cours de validation)</p>
--	---

➤ **Historique des versions**

Date	N° de version	Statuts	Auteur(s)	Affiliation de l'auteur	Validé par	Validé le
16/02/2024	1	En cours	Marie GABORIT Annabelle BACHELET	UEP 0326 – Unité Expérimentale du Pin	Frédéric LAUNAY (Directeur)	27/06/2024
11/07/2024	2	Envoyé pour relecture	Marie GABORIT Annabelle BACHELET	UEP 0326 – Unité Expérimentale du Pin	Frédéric LAUNAY	12/07/2024
30/08/2024	3	Envoyé en validation	Marie GABORIT Annabelle BACHELET	UEP 0326 – Unité Expérimentale du Pin	Frédéric LAUNAY (Directeur)	30/08/2024

➤ **Contributeurs**

Nom Prénom	Affiliation	Rôles
BACHELET Annabelle	INRAE	Personne contact pour les données (données zootechnique) <a href="https://orcid.org/0009-0002-6863-0829">https://orcid.org/0009-0002-6863-0829</a>
GABORIT Marie	INRAE	Personne contact pour les données (données zootechnique) <a href="https://orcid.org/0009-0008-5461-1196">https://orcid.org/0009-0008-5461-1196</a>
HERAULT Frédéric	INRAE	RDO du département GA <a href="https://orcid.org/0000-0001-5824-4081">https://orcid.org/0000-0001-5824-4081</a>
		Coordinateur du projet
LAUNAY Frédéric	INRAE	Personne contact pour les données (environnement)
		Responsable du plan de gestion de données <a href="https://orcid.org/0000-0001-5358-9400">https://orcid.org/0000-0001-5358-9400</a>

## Objectifs du Plan de Gestion de Données (PGD)

Le plan de gestion de données de l'unité est un document cadre décrivant le cycle de vie des données produites ou utilisées dans le collectif.

Il se décline en 2 produits de recherche regroupant l'ensemble des typologies de données produites ou mobilisées au sein de l'unité. Ces deux produits de recherches concernent :

- Les données issues de mesures ou observations effectuées sur les animaux
- Les données issues de mesures ou de relevés de l'environnement d'élevage

Il pourra servir de référence pour l'élaboration de plan de gestion de données de projet de recherche s'appuyant sur les travaux menés dans l'unité.

PGD en lien avec les PGD :

- Plan de gestion de structure du Cati SICPA V1 20230830 → doi:10.57745/ACTK8H
- PGD – AgroClim (encours de rédaction)

## Produits de recherche décrits dans ce PGD

1. Ensemble des données issues de mesures ou observations effectuées sur les animaux
2. Ensemble des données issues de mesures ou de relevés de l'environnement d'élevage

## III- Présentation Générale des données

- 1) Ensemble des données issues de mesures ou observations effectuées sur les animaux

### *Mode d'obtention des données*

- Données générées par la structure
- Données produites par un tiers

Les données de phénotypage animal sont collectées pendant toute la vie des animaux.

**Données générées par la structure issues des observations visuelles et mesures instrumentées:** Ces données sont collectées via des saisies papiers, des saisies informatiques sur tableurs ou l'utilisation de SICPA. (cf PGD Cati SICPA). Les données sont stockées sous différents formats: papiers, tableurs Excel (.xlsx ou .csv) ou format texte (.txt; .docx; .pdf)

**Données produites par un tiers:** les prélèvements de types métaboliques et composition du lait sont faits sur site mais ils sont par la suite envoyés à des laboratoires extérieurs indépendants qui enverront le rapport d'analyse avec les données obtenues (format pdf, excel, txt) qui seront ensuite intégrées dans le système d'information SICPA Bovin, SICPA alimentation et SICPA Sanitaire.

**Les données issues de capteurs** (accéléromètres, bolus pH) mais aussi celles des GreenFeed et SNIFFER (méthane) sont récupérables via des interfaces web. Les échantillons d'aliments en secs ou en congelés (fourrages, concentrés) sont envoyés à des laboratoires privés (LANO ou LABOCEA) et les résultats

nous sont envoyés en fichiers informatiques (.xlsx, /pdf). Ces données sont enregistrées sur le DATA Center dans un fichier spécifique.

Le dispositif de scan NEC (Note d'état corporel) 3D des bovins fait l'objet d'acquisition d'images qui sont envoyées à l'entreprise l'ayant développé qui les analyse et les reconstitue en 3D.

**\*\*Acronymes\*\***

SICPA → Système d'Information et Calcul pour le Phénotypage Animal

### Origine

- Analyse
- Corpus audiovisuel
- Expérimentation
- Observation

Les observations sont obtenues par relevés visuels ou toucher des animaux.

Les mesures instrumentées sont obtenues avec au moins un outil de phénotypage lors d'expérimentations :

<b>Méthode d'enregistrement</b>	<b>Données obtenues</b>	<b>Format des données</b>
<i>Caméras 3D Ouest</i>	NEC automatisées – images	<b>Format jpeg, png</b>
<i>GreenFeed et Sniffer</i>	CH4 – données numériques	<b>Fichier xlsx ou csv</b>
<i>Capteurs Medria</i>	Comportement – données numériques ou binaires	<b>Fichier xlsx ou csv</b>
<i>Digestibilité ovins</i>	Digestibilité des fourrages – données numériques	<b>Fichier xlsx ou csv</b>
<i>Echantillonnages des fourrages et aliments</i>	Matière sèche des aliments – données numériques	<b>Fichier xlsx ou csv</b>
<i>Laboratoires extérieurs (LABOCEA, LANO) – INRAtion V5</i>	Valeurs des aliments produits sur l'unité	<b>Fichier xlsx ou csv</b>
<i>OptiGen (logiciel d'accouplement) retranscrit sur Excel</i>	Plan d'accouplement – données texte	<b>Fichier xlsx ou csv</b>
<i>Bon de livraison des doses, feuille de suivi des consommations de doses</i>	Stock cuve (doses IA, embryons) données numériques et texte	<b>Fichier pdf</b> <b>Fichier xlsx ou csv</b>
<i>Génotypage (cartilage ou prise de sang)</i>	Index des animaux – données numériques	<b>Fichier xlsx ou csv</b>
<i>Herbomètre, enregistrement Excel des pratiques de pâturage</i>	Suivi du pâturage – données numériques et texte	<b>Fichier xlsx ou csv</b>
<i>Compte rendu de visite ostéopathe</i>	Suivi ostéopathique – données texte	<b>Fichier pdf</b>
<i>Pièges photographiques horodatées et fichier Excel partagé</i>	Heures d'entrée et de sortie de la parcelle pâturée – images et données numériques	<b>Format jpeg, png</b> <b>Fichier xlsx ou csv</b>
<i>Compteurs à lait salle de traite et robot de traite puis saisie dans SICPA Bovins</i>	Production laitière – données numériques	<b>Fichier xlsx ou csv</b>



<i>Echantillonneurs de l'unité puis analyses laboratoire extérieur</i>	Composition du lait (TB, TP, CCS, Spectres MIR, Acétone, Urée) – données numériques	<b>Fichier xlsx ou csv</b>
<i>Bascules unité puis saisie dans SICPA Bovins</i>	Poids des animaux – données numériques	<b>CATI SICPA</b>
<i>Agents notateurs formés sur l'unité</i>	NEC des animaux – données numériques	<b>CATI SICPA</b>
<i>Auges Biocontrol, pesée dessileuse, pesée refus puis saisie dans SICPA Alimentation</i>	Quantités d'aliments ingérées – données numériques	<b>CATI SICPA</b>
<i>Saisie papier puis SICPA Sanitaire</i>	Evènements sanitaires – données numériques et texte	<b>CATI SICPA</b>
<i>SICPA Bovins</i>	Localisation des animaux – données texte	<b>CATI SICPA</b>
<i>Saisie papier puis BDD d'identification nationale et SICPA Bovins</i>	Etat civil des bovins – donnée texte	<b>CATI SICPA</b>
<i>Mesures réalisées par agents puis saisie dans SICPA Bovins</i>	Mensurations génisses et vaches laitières – données numériques	<b>CATI SICPA</b>
<i>Pointage (note de la morphologie des bovins) réalisé par technicien OS</i>	Pointage vaches laitières en première lactation – données numériques	<b>Document papier Fichier xlsx ou csv</b>

**\*\*Acronymes utilisés\*\***

SICPA → Système d'Information et Calcul pour le Phénotypage Animal

Les données transformées (analyse) correspondent quant à elles à des données relevées par observation ou de manière instrumentée et qu'il a fallu regrouper et analyser avant leur diffusion. Les données d'alimentation par exemple sont relevées, enregistrées dans un tableur et modifiées en fonction des besoins de l'expérimentation avant diffusion.

Enregistrement / **Corpus Audiovisuel** → enregistrement d'images

**\*\*Acronymes\*\***

IA → insémination artificielle

BDD → base de données

OS → organisme de sélection

NEC → note d'état corporel

TB → taux butyrique

TP → taux protéique

CCS → comptage de cellules somatiques

### *Type de données*

- Audiovisuel
- Dataset
- Image

Prépondérance des données de type dataset

**Observations** → NEC, BEA, tests comportements

**Instrumentées** → différentes pesées, échographie, colliers medria, données alimentation, notation carcasse avec enregistrement informatisé

**Transformées** → données alimentation des bovins

**Prélèvements** → Manuel

**Prélèvements analysés à l'extérieur de l'unité** → analyses contrôle laitier, analyses sanguines, coproscopies, analyses jus de rumen, analyses photos

**Traçabilité** → Données sanitaire

**\*\*Acronymes\*\***

NEC → Note État Corporel

BEA → Bien-être Animal

### *Nature des données*

**Données de mensurations des animaux** : Poids (pesée ponctuelle des animaux enregistrée en gramme), NEC (notation de 0 = extrêmement émacié à 5 = vache très grasse - enregistrement de la notation au dixième),

**Données physiologiques** : BEA (notation d'absence ou présence d'anomalie pour chaque critère d'une grille d'évaluation en fonction de l'expérimentation), échographie

**Données de parasitologie** : coproscopies, fécès

**Données sanitaires** : ensemble des ordonnances reçues, traitements réalisées ou évènements sanitaire qui peuvent être faits

**Données d'élevage** : contrôle laitier (qualité du lait = TB et TP (grammes/kilogrammes) ; quantité d'urée et de lactose en milligrammes par litre ; dénombrement du nombre de cellules somatiques et quantité de lait mesuré en litres par vache par jour) ; données d'alimentation par la mesure de l'ingestion individuelle (enregistrement des quantités distribuées en kilo par vache par distribution + enregistrement des quantités non consommées = refus en kilo par vache par jour ; enregistrement individuel ou par lot) ; Données alimentation en suivi quotidien (alimentation du troupeau avec quantité en kilogrammes par jour distribuée par lot)

**Données des accéléromètres portés par les animaux (colliers)**

**\*\*Acronymes utilisés\*\***

NEC → Note d'Etat Corporel

BEA → Bien-Etre Animal

SICPA → Système d'Information et Calcul pour le Phénotypage Animal

### *Format des données*

**Voir tableau page 6**

### *Périmètre thématique des données*

- Elevage et Productions Animales
- Santé Animale et Pathologie
- Systèmes Agricoles et pratiques d'élevage

## 2) Ensemble des données issues de mesures ou de relevés de l'environnement d'élevage

### *Mode d'obtention des données*

- Données générées par la structure
- Données produites par un tiers

**Données générées par la structure** : données issues des observations visuelles des végétations/cultures et mesures instrumentées (station météo), enregistrement des rendements et des opérations culturales. Ces données sont collectées via des saisies papiers, des saisis tableurs ou l'utilisation de SICPA Alimentation. Les données sont stockées sous différents formats : papiers, tableurs Excel (.xlsx ou .csv) ou format texte (.txt; .docx; .pdf)

**Données produites par un tiers** : données issues des échantillons de sol, de lisier et de fumier qui sont analysés dans un laboratoire extérieur.

**\*\*Acronymes\*\***

SICPA → Système d'Information et Calcul pour le Phénotypage Animal

### *Origine*

- Observation
- Enregistrements

Les observations sont obtenues par relevés visuels des végétations pour le suivi de l'évolution des prairies.

Les interventions sur les parcelles sont enregistrées directement sur MesParcelles (outil web des Chambres d'Agriculture : <https://mesparcelles.fr/> )

Les mesures instrumentées sont obtenues avec au moins un outil de phénotypage :

La station météorologique : sondes... → données météorologiques

Capteurs thermiques → mesures d'ambiance

Les données transformées (analyse) correspondent quant à elles à des données relevées par observation ou de manière instrumentée et qu'il a fallu regrouper et analyser avant leur diffusion.

## *Type de données*

- Dataset

Observations → relevés de végétation

Enregistrements instrumentés → sondes de températures et d'humidité via la station météo du domaine, données cultures : rendements, date de semis/récolte, données d'ambiance via des capteurs en bâtiment

## *Nature des données*

**Données météorologiques** : date et heure de la mesure ; température en degré Celsius (T) ; Humidité instantanée en % (U) ; Hauteur des précipitations en millimètres (RR) ; Rayonnement en Joules par centimètre carré (PAR\_H) ; Rayonnement global en joules par centimètre carré (RG) ; Vitesse maximale du vent en mètre par seconde (VX)

**Données d'ambiance** : capteurs BKer Température en degrés Celsius ; Hygrométrie en % correspondant au taux d'humidité

**Données culturales** : date des interventions : semis, récolte, épandages, quantité récoltée en Tonnes de matière brute, nom et surface de parcelle concernée, composition de la parcelle ; échantillon de chaque fourrage → % de matière sèche ; MM + MAT + CB + NDF + ADF + ADL en grammes par kilogrammes de matière sèche ; SMS + SMO en % de MO

**Données végétation** : suivi du nombre de jours de pâturage par parcelle ; relevés de biomasse en kilogrammes par hectare et de Matière Sèche en %

**Données sol et effluents d'élevage** : l'UE réalise tous les ans des prélèvements de sol, de lisier et de fumier. Ces échantillons sont envoyés dans un laboratoire indépendant et nous récupérons les données issues de ces analyses.

### **\*\*Acronymes\*\***

MM → Matières Minérales

MAT → Matière Azotée Totale

CB → Cellulose Brute

NDF → Neutral Detergent Fiber

ADF → Acid Detergent Fiber

ADL → Acid Detergent Lignin

SMS → Solubilité ou dégradabilité enzymatique de la Matière Sèche

SMO → Solubilité ou dégradabilité enzymatique de la Matière Organique

MO → Matière Organique

*Format des données*

<b>Origine des données</b>	<b>Format des données</b>
<i>Données culturales</i>	Saisie en ligne sur le site web MesParcelles
<i>Données de pâturage</i>	Tableur Excel
<i>Données météorologiques</i>	PGD Agroclim = Application Climatik ( <a href="https://agroclim.inrae.fr/">https://agroclim.inrae.fr/</a> )
<i>Données ambiance bâtiment</i>	Fichier xlsx ou csv
<i>Données végétation</i>	Fichier xlsx ou csv
<i>Données sols et effluents d'élevage</i>	Fichier pdf
<i>Résultats d'analyse des fourrages</i>	Fichier xlsx ou csv

**\*\*Acronymes utilisés\*\***

SICPA → Système d'Information et Calcul pour le Phénotypage Animal

*Périmètre thématique des données*

- Elevage et Productions Animales
- Santé Animale et Pathologie
- Systèmes Agricoles et pratiques d'élevage
- Agriculture et productions végétales
- Santé Végétale et Pathologie
- Sols et sciences du sol

## IV- Droits de propriété intellectuelle

*Qui détiendra les droits sur les données et les autres informations créées ?*

INRAE possède les droits de propriété sur les données brutes produites par l'unité en dehors des projets.

Dans le cadre des projets, la propriété des données est déterminée dans un accord de consortium et décrite dans le PGD de projet.

## V- Sensibilité des données

*Identification du niveau de sensibilité des jeux de données (pouvant être utilisé pour identifier une personne)*

→ Diffusion limitée

Les données sensibles correspondent aux :

- Identifiants uniques des animaux
- Noms ou images des personnes réalisant les mesures (données à caractères personnelles)

Leur accès libre sera soumis à autorisation, suite à une éventuelle anonymisation.

*Quelles sont les mesures prises et les normes auxquelles il est nécessaire de se conformer pour garantir la sécurité des données sensibles ?*

Dans le cadre de la confidentialité liée à un projet, la sécurité sera déterminée par le PGD du projet.

Dans le cadre des autres jeux de données, le contrôle de l'accès aux données est actuellement limité aux membres de l'UEP et aux partenaires scientifiques identifiés (PGD CATI SICPA).

Les données stockées à l'extérieur de l'unité (conditions générales d'utilisation des GreenFeed avec C-Lock (données de méthane) et contrat (C4921) Medria Solutions (données comportementales)) sont régies par les termes des CGU ou du contrat.

Pour les données internes non stockées dans le CATI SICPA, elles sont stockées dans les espaces partagés de l'UEP, dont l'accès est limité. L'ensemble des agents est informé des restrictions de diffusion de ces informations.

*S'il y a des données à caractère personnel, quelles sont les mesures envisagées pour les protéger au cours du projet ou dans le cadre d'une réutilisation ?*

Les seules données à caractère personnel sont celles des agents INRAE concernés par les protocoles et les expérimentations. Le nom des agents est anonymisé lors des exports des données. L'accès à ces informations non anonymisées est soumis à autorisation.

## VI- Partage des données

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

L'Unité Expérimentale du Pin dépend d'un organisme de recherche public (INRAE), les données qu'elle produit doivent être communicables et accessibles à toute personne qui en fait la demande si elles n'entrent pas dans le cadre des exceptions légales (données confidentielles).

*Quelles sont les réutilisations potentielles de ces données ?*

- Utilisation dans la modélisation de système
- Publications scientifiques
- Vulgarisation et diffusion scientifiques
- Travaux de Recherche et Développement
- Enseignement

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

La lecture des données d'élevage conservées dans les systèmes d'information du CATI SICPA peut se faire avec le logiciel de développement du même nom (Voir CATI SICPA)

- Pour les fichiers numériques : les logiciels de la suite Office sont utilisés, notamment Excel et Word. Pour la consultation la plupart des logiciels de traitement de texte et tableur permettent la lecture des fichiers. De même, l'utilisation d'un navigateur web classique permet d'extraire les données stockées sur les serveurs des sociétés prestataires.

- Pour les versions papier : la consultation sur site est possible par tous les agents de l'UEP

*Comment les données seront-elles partagées ?*

- Envoi par e-mail des données et/ou des fichiers partenaires : lien FileSender
- Espace partagé en ligne : espaces partagés
- CATI Sicpa

*Avec qui ?*

- Partenaires identifiés
  - Aux agents INRAE authentifiés (LDAP) qui ont les autorisations
  - Aux partenaires identifiés suivant les projets de recherche

*Sous quelle licence ?*

- Licence ouverte <https://www.etalab.gouv.fr/licence-ouverte-open-licence> (compatible CC-BY)

## VII- Organisation et documentation des données

*Quels méthodes et outils sont utilisés pour acquérir et traiter les données, depuis leur acquisition jusqu'à leur mise à disposition, leur archivage ou leur destruction ? Utiliser éventuellement un lien vers un schéma illustrant les processus.*

- Collectées à l'aide :
  - Papier – Crayon
  - PDA
  - Appareil photo, caméras, drones
  - Sondes
  - Prélèvements
- Traitées sur :
  - Fichier informatique tableur Excel
  - CF PGD Sicpa Bovin
  - CF PGD INRAE UE Agroclim
- Stockées sur :
  - Sur le Datacenter INRAE puis sauvegardées quotidiennement sur le serveur UEP (NAS)
  - Stockage local individuel (clé USB, disque dur PC)
  - Classeur (données format papier)
  - Sharepoint, Nextcloud, espaces partagés

<i>Nature / type de données</i>	<i>Sauvegarde des données</i>
<i>NEC automatisées - images</i>	Sur serveur du prestataire
<i>CH4 - données numériques</i>	Sur serveur du prestataire
<i>Comportement - données numériques ou binaires</i>	Sur serveur du prestataire
<i>Digestibilité des fourrages - données numériques</i>	Sur Datacenter/NAS local
<i>Matières sèches des aliments - données numériques</i>	Sur Datacenter/NAS local
<i>Valeurs des aliments produits sur l'unité - données numériques</i>	Sur Datacenter/NAS local
<i>Plan d'accouplement - données texte</i>	Sur Datacenter/NAS local

<i>Stock cuve (doses IA, embryons) - données numériques</i>	Sur Datacenter/NAS local
<i>Index animaux - données numériques</i>	BDD CTIG
<i>Suivi du pâturage - données numériques et texte</i>	Sur Datacenter/NAS local
<i>Suivi ostéopathique - données texte</i>	BDD CTIG
<i>Heures d'entrée dans parcelle pâturées - images</i>	Sur Datacenter/NAS local
<i>Météologie - données numériques</i>	Sur Datacenter/NAS local
<i>Production laitière - données numériques</i>	BDD CTIG
<i>Composition du lait - données numériques</i>	BDD CTIG
<i>Poids des animaux - données numériques</i>	BDD CTIG
<i>Notes d'état corporel - données numériques</i>	BDD CTIG
<i>Quantités d'aliments ingérées - données numériques</i>	BDD CTIG
<i>Événements sanitaires - données texte</i>	BDD CTIG
<i>Localisation des animaux - données texte</i>	BDD CTIG
<i>État civil des bovins - données texte</i>	BDD CTIG
<i>Mensurations génisses et vaches laitières - données numériques</i>	BDD CTIG
<i>Pointage vaches laitières première lactation - données numériques</i>	BDD CTIG

<i>Etape</i>	<i>Méthode(s)</i>	<i>Outil(s)</i>
<b>Préparation</b>	Rédaction de protocoles selon la trame qualité (E-SMQ 18 - Modèle protocole expérimental)	Traitement Texte (Word)
<b>Acquisition</b>	Procédures, protocoles et modes opératoires	- Saisie Papier - Outils de mesures ou enregistrement de boucles (balance, bâton...)
<b>Traitement de données</b>	pré-traitement des données, tris des données aberrantes	- CF PGD SICPA - tableur (Excel)
<b>Mise à disposition</b>	- envoi par mail - mise en ligne sur espace de partage - logiciel SICPA	- CF PGD SICPA - mail / espace partagé
<b>Archivage</b>	Non défini aujourd'hui	
<b>Destruction</b>	Non défini aujourd'hui	

*Quelles métadonnées seront utilisées pour accompagner le jeu de données ? Quels seront les standards, vocabulaires, taxonomies... utilisés pour décrire et représenter les données et éléments de métadonnées ? Comment les métadonnées seront-elles produites et mises à jour ?*

Les données produites font principalement référence aux ontologies suivantes :

- ATOL = caractéristiques phénotypiques (<https://agroportal.lirmm.fr/ontologies/ATOL>)
- AHOL = caractéristiques liées à la santé animale (<https://agroportal.lirmm.fr/ontologies/AHOL>)
- EOL = caractéristiques liées à l'environnement (<https://agroportal.lirmm.fr/ontologies/EOL>)

ainsi qu'au Thésaurus INRAE : <https://thesaurus.inrae.fr/thesaurus-inrae/fr/>



*Une documentation complémentaire aux métadonnées est-elle nécessaire pour décrire les données et assurer leur réutilisabilité sur le long terme ?*

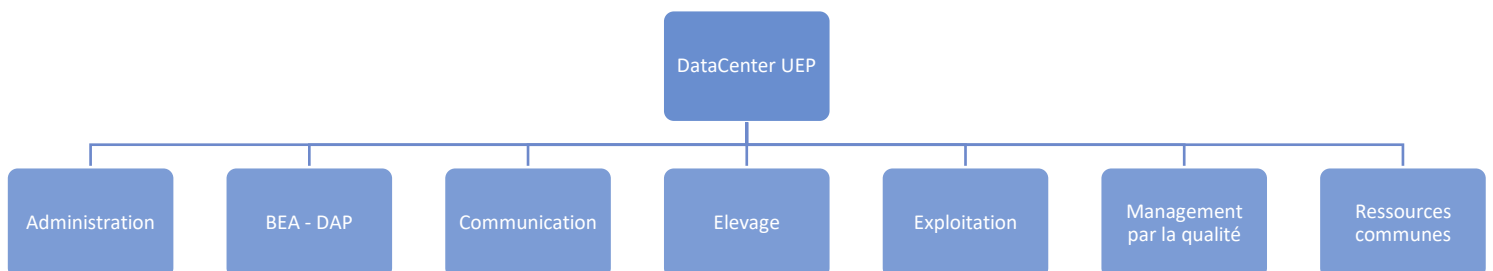
- Les protocoles expérimentaux ne sont pas décrits dans les SICPA espèces. Bien que l'on puisse identifier si les animaux étaient inclus ou non dans des protocoles expérimentaux lors de l'acquisition des données, on ne dispose que de l'identifiant du protocole et des lots.
- Le SICPA sanitaire et alimentation se connecte aux SICPA espèces pour récupérer les informations concernant le suivi des animaux (animaux, présents, stade, âge, localisation, mise en expérimentation, ...)
- Concernant les données d'alimentation, il existe un fichier de suivi de l'alimentation dans lequel sont inscrits les éléments modifiés sur les données brutes avant leur importation dans le SICPA.
- Pour les données acquises dans le cadre de projets scientifiques, c'est le PGD de ces derniers qui fait foi en matière de métadonnées

*Comment les fichiers de données sont-ils gérés et organisés : contrôle des versions, conventions de nommage des fichiers, organisation des fichiers*

Cf PGD Cati SICPA pour les données saisies sous les SI

Une réflexion globale est en cours pour organiser le stockage des fichiers sur le DATACENTER.

L'arborescence principale est la suivante et les niveaux d'accès en lecture et/ou modification sont gérés avec les LDAP des agents/chercheurs travaillant avec l'UEP. L'accès le plus restreint concerne évidemment la partie « administration » uniquement accessible aux DU, DUA et GU(s). Les autres dossiers sont tous accessibles « en mode lecture » et certains « en modification » pour faciliter la communication, l'enregistrement et la transmission des données.



*Quel est le processus de contrôle qualité des données ?*

La donnée produite :

- 1- calibration et étalonnage des outils de mesure, formation des agents à l'utilisation et la manipulation de ces outils
- 2- données faisant appel à une expertise (ex : Note d'Etat Corporel (NEC)) les agents sont formés et il y a un processus de contrôle interne (double notation)
- 3- acquisition de la donnée et vérification technique par le collecteur
- 4- validation scientifique des données au fil de l'eau et une fois par an pour les bilans
- 5- vérification par le SI (cf PGD CATI Sicpa)

## Stockage et sécurité des données

*Les systèmes d'information de la structure ont-ils fait l'objet d'une analyse de risques ou d'une homologation ?*

- Non

cf PGD du CATI SICPA

Pour les données hors CATI, il n'y a pas d'analyse de risque ou d'homologation de l'ensemble des systèmes informatiques sur lesquels elles sont stockées.

Néanmoins, la plupart des données sont stockées sur les espaces partagés du serveur de l'unité qui n'a pas fait l'objet d'une analyse de risque.

*Quels types de supports physiques sont utilisés pour stocker les données ?*

- Stockage des données sur les SICPA
- Stockage des données sur le réseau partagé UEP (serveur local NAS)
  - Stockage ponctuel sur ordinateur, clé ou disque dur
  - Archivage des agendas et cahiers pour les événements d'élevage ou sanitaires notés d'abord sur papier
- Stockage des données sur un dossier partagé (DataCenter INRAE)

*Quelles sont les mesures de sécurité mises en place lors des étapes de transfert des données ?*

Le transfert des données se fait par FileSender ou via le réseau INRAE sous réserve d'authentification par le LDAP.

Pour les données sur le SICPA, cf PGD SICPA.

*Quelle est la volumétrie actuelle et prévisionnelle ?*

La volumétrie actuelle sur le partage est de 15 Go

*L'entité hébergeant physiquement les données a-t-elle une politique de sécurité de l'information et a-t-elle un plan d'assurance sécurité ?*

Ce sont les politiques du Data center et de SICPA qui gèrent ces aspects sécurité.

Le partage hébergé sur le DataCenter INRAE est répliqué afin d'assurer la sécurité de l'information.

Les données stockées sur les SI (bovins, sanitaire et alimentation) du SICPA sont sur des bases de données géré par le CTIG qui en assure la sécurité.

*Sécurité - Confidentialité : les données font-elles l'objet d'échange ou de partage avec de tiers acteurs et selon quelles modalités ? comment sont déterminés les droits d'accès aux données avant leur publication ?*

Certaines données sont transférées aux partenaires chercheurs, qui assurent les droits d'accès à ces données jusqu'à la publication.

Les partenaires (techniciens, chercheurs) souhaitant obtenir des données doivent faire la demande auprès de L'UE Le Pin

*Sécurité - Intégrité - Traçabilité : Quelles sont les mesures de protection mises en œuvre pour suivre la production et l'analyse des données ?*

Cahiers de laboratoire utilisés par les équipes des deux troupeaux pour noter tous les évènements ou commentaires liés à la production de données qui pourraient faire varier sa valeur ou son interprétation

- Horodatage des données lors de l'utilisation d'outils numériques ou inscription de la date lors de relevés manuels
- Un système de copie automatique quotidienne des données enregistrées sur chaque PC de l'unité est réalisé sur le serveur de l'unité.

*Les agents de la structure ont-ils bénéficié d'une sensibilisation aux bonnes pratiques d'hygiène numérique ?*

- Oui

## VIII- Archivage et conservation des données

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

Toutes les données de phénotypage sont conservées à long terme.

Cf PGD Cati SICPA

*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 ?*

Cf PGD Cati SICPA

*Quelle est la durée de conservation des données ?*

Cf PGD Cati SICPA

Pour les données hors CATI, il n'y a pas de durée de conservation définie. Par défaut, illimité.

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

Cf PGD Cati SICPA

Pour les données hors CATI, il n'y a pas de politique de financement du stockage définie à ce jour.