



**HAL**  
open science

## Plan de gestion de données de la structure Plateforme Tropicale d'Expérimentation sur l'Animal

Nathalie Minatchy, Frédéric Herault, Madly Moutoussamy

### ► To cite this version:

Nathalie Minatchy, Frédéric Herault, Madly Moutoussamy. Plan de gestion de données de la structure Plateforme Tropicale d'Expérimentation sur l'Animal. INRAE. 2024, 25 p. hal-04854357

**HAL Id: hal-04854357**

**<https://hal.inrae.fr/hal-04854357v1>**

Submitted on 23 Dec 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 - ShareAlike 4.0 International License



 Centre  
Antilles-Guyane

**INRAE**

1071743  
PLATEFORME TROPICALE  
D'EXPÉRIMENTATION SUR L'ANIMAL [PTEA]



## Plan de Gestion de Données de la structure "Plateforme Tropicale d'Expérimentation sur l'Animal"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle  
"INRAE - Modèle Structure"



## **PGD DE LA STRUCTURE**

### **"PLATEFORME TROPICALE D'EXPÉRIMENTATION SUR L'ANIMAL"**

*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.*

### **ACRONYMES ET ABBRÉVIATIONS UTILISÉS**

ADF → Acid Detergent Fiber

ADL → Acid Detergent Lignin

CATI → Centre automatisé de traitement de l'information

CB → Cellulose Brute

DAC → Distributeur automatique de concentré

MAT → Matière Azotée Totale

MM → Matières Minérales

MO → Matière Organique

NDF → Neutral Detergent Fiber

NEC → Note d'Etat Corporel

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

SMO → Solubilité ou dégradabilité enzymatique de la Matière Organique

SMS → Solubilité ou dégradabilité enzymatique de la Matière Sèche

## RENSEIGNEMENTS SUR LE PLAN

Titre du plan	PGD de la structure "Plateforme Tropicale d'Expérimentation sur l'Animal"
Livrable	PGDPTEA_V1
Version	Version initiale
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.</p>
Domaines de recherche (selon classification de l'OCDE)	Agriculture, forestry, and fisheries
Langue	fra
Date de création	2022-11-30
Date de dernière modification	2024-09-04
Identifiant	attendre n° DOI du plan
Type d'identifiant	DOI
Licence	Nom Creative Commons Attribution Share Alike 4.0 International URL <a href="https://creativecommons.org/licenses/by-sa/4.0/">https://creativecommons.org/licenses/by-sa/4.0/</a>
Documents (publications, rapports, brevets, plan expérimental....), sites web associés	Site web : <a href="https://ptea.antilles.hub.inrae.fr/">https://ptea.antilles.hub.inrae.fr/</a>
Plans de gestion en lien avec le projet ou l'activité de l'entité	<ul style="list-style-type: none"> <li>• PGD du CATI SICPA : <a href="https://doi.org/10.57745/ACTK8H">https://doi.org/10.57745/ACTK8H</a></li> <li>• PGD de l'Unité de recherches Agroécologie, génétique et systèmes d'élevage tropicaux (UR ASSET)</li> </ul>

## RENSEIGNEMENTS SUR L'ENTITE

Nom de la structure	Plateforme Tropicale d'Expérimentation sur l'Animal
Acronyme	PTEA
Résumé	<p>La Plateforme Tropicale d'Expérimentation sur l'Animal (PTEA) met en place des expérimentations sur les animaux d'élevage en milieu tropical humide. Elle est localisée en Guadeloupe sur deux sites (Duclos : 15 ha et Gardel : 50 ha) offrant des situations pédoclimatiques contrastées (pluviométrie, nature des sols). Afin de favoriser la transition agroécologique des systèmes polyculture-élevage insulaires tropicaux, la PTEA offre son appui à la recherche pour contribuer à :</p> <ul style="list-style-type: none"> <li>• La production de connaissances sur le phénotypage animal en milieu tropical, sur l'évaluation multicritère des ressources alimentaires et des systèmes d'élevage,</li> <li>• La conception et l'évaluation d'innovations biotechniques, organisationnelles et logicielles pour favoriser la transition agroécologique des systèmes polyculture-élevage insulaires tropicaux,</li> <li>• La conservation de ressources génétiques animales locales (races Créole) en association avec le CRB CARARE.</li> </ul>
Sources de financement	<ul style="list-style-type: none"> <li>• INRAE</li> <li>• Ventes de produits</li> <li>• Projets nationaux ou européens</li> </ul>
Date de début	30/11/2022
Date de fin	31/12/2028
Partenaires	<ul style="list-style-type: none"> <li>• Unité de recherches Agroécologie, génétique et systèmes d'élevage tropicaux (UR ASSET)</li> <li>• Département de Génétique Animale (GA)</li> <li>• Autres unités INRAE du département GA</li> <li>• CIRAD</li> <li>• Institut technique de l'élevage (ITEL)</li> <li>• Institut technique tropical (IT2)</li> <li>• Établissement public local d'enseignement et de formation professionnelle agricoles (EPLEFPA)</li> <li>• Interprofession élevage (IGUAVIE)</li> </ul>

**Produits de recherche :**

1. Ensemble des données zootechniques et expérimentales recueillies sur l'animal (Jeu de données)
2. Ensemble de données liées à l'environnement (biomasses, composition d'aliments, température, hygrométrie...)(Jeu de données)

**Contributeurs**

<b>Nom</b>	<b>Affiliation</b>	<b>Rôles</b>
MINATCHY Nathalie	INRAE	<ul style="list-style-type: none"><li>• Coordinateur du projet</li><li>• Personne contact pour les données (Animaux, Environnement)</li><li>• Responsable du plan de gestion de données</li></ul>

# PGD DE LA STRUCTURE "PLATEFORME TROPICALE D'EXPÉRIMENTATION SUR L'ANIMAL"

## INFORMATIONS SUR LA STRUCTURE

### Nom de la structure

UE1294 PTEA Plateforme Tropicale d'Expérimentation Animale (changement de nom au 01 janvier 2025 : Plateforme Tropicale Elevage pour l'Agroécologie)

### Type de structure

- Unité Expérimentale

*IDENTIFIANT DE LA STRUCTURE PRÉCISER LE FOURNISSEUR DE L'IDENTIFIANT (ISNI, VIAF, FUNDREF, DATA CITE...).*  
200817784A (RSNR)

### RESPONSABILITES DANS LA STRUCTURE

Nom, Prénom	Courriel	Rôle
MINATCHY Nathalie	nathalie.minatchy@inrae.fr	Directrice d'unité, Responsable du PGD, Gestionnaire des données, site de Duclos
ARQUET Rémy	remy.arquet@inrae.fr	Directeur-adjoint, Gestionnaire des données, site de Gardel
GIORGI Mario	mario.giorgi@inrae.fr	Directeur-adjoint, Responsable de la qualité des données

### Etablissement(s) tutelle(s)

INRAE, Institut national de la recherche pour l'agriculture, l'alimentation et l'environnement.

### Département de rattachement INRAE (ou anciens départements Inra)

- GA : Génétique animale

### Financier(s) (permettant l'acquisition des jeux de données - hors projet)

- INRAE,
- Ventes de produits et de prestations,
- Subventions régionales via le FEDER (Fonds européen de développement Régional),
- Aides de soutien à l'agriculture apportée par l'État via la Direction de l'alimentation, de l'agriculture et de la forêt.

## INFORMATIONS SUR LE PLAN DE GESTION

DOI (VERSION PUBLIEE DU PLAN DE GESTION)

A venir après publication sous HAL

### Historique des versions

Date	n° de version	Statuts	Auteurs	Affiliation de l'auteur (se reporter à l' <a href="#">annuaire INRAE</a> )	Validé par	Validé le
19/08/2024	1	en cours	Nathalie MINATCHY	UE 1294 PTEA		
			Frédéric HERAULT	UMR PEGASE		

## PRÉSENTATION GÉNÉRALE DES DONNÉES

### ENSEMBLE DES DONNÉES ZOOTECHNIQUES ET EXPÉRIMENTALES RECUEILLIES SUR L'ANIMAL

#### MODE D'OBTENTION DES DONNEES

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

La PTEA est une unité expérimentale INRAE possédant la plupart des espèces de rente : porcs, petits-ruminants (chèvres et moutons), bovin et volaille. Elle a vocation à permettre l'expérimentation sur des questions de recherche et des développements méthodologiques et technologiques sur l'élevage. Une partie des données est générée en routine (exemples : pesées régulières des animaux, actes de soin, vétérinaires etc...). Une autre partie des données est liée aux expérimentations et possède un caractère plus ponctuel. L'unité permet de générer des données brutes, c'est à dire produites par observation ou expérimentation. Certaines données brutes doivent être traitées avant diffusion (données transformées). D'autres données sont générées par des tiers (par exemple laboratoires d'analyse pour traiter les échantillons sanguins ou de fèces).

- **Données générées par la structure :**

Les données générées par la structure sont issues d'observations (visuelles ou par manipulation des animaux), ou de mesures instrumentées sur les animaux.

Les données instrumentées sont généralement obtenues à l'aide d'un outil de phénotypage (par exemple des balances) et peuvent-être collectées en routine ou ponctuellement pour répondre à des besoins expérimentaux.

Les données sont collectées via des saisies papiers, des saisies Excel ou l'utilisation des systèmes d'information, notamment les systèmes d'information et de calcul pour le phénotypage animal (SICPA) qui sont décrits dans le PGD du CATI SICPA (<https://doi.org/10.57745/ACTK8H>).

Les données sont ensuite stockées sous différents formats : papier, tableur (.xlsx ou .csv) ou format texte (.txt; .docx; .pdf).

Le stockage des données collectées via les outils du CATI SICPA, se fait dans des datacenters INRAE (Région parisienne et Toulouse), sous forme de bases de données relationnelles ou NoSQL.

- **Données produites par un tiers :**

Il s'agit ici de données produites à partir de prélèvements de tissus biologiques, de sang, de fèces, ou d'échantillons d'aliments qui sont réalisés sur site par du personnel habilité à l'expérimentation animale. Les échantillons sont ensuite envoyés pour analyse aux laboratoires.

Les rapports d'analyses et les données produites sont réceptionnés au format .pdf, .xlsx, .csv, .tab, txt.

## Origine

- Analyse
- Corpus audiovisuel
- Expérimentation
- Observation

Les observations, obtenues par relevés visuels ou toucher des animaux, sont les suivantes :

- Suivi zootechnique (reproduction, parasitisme, pointage) : Notes d'état corporel (NEC), score FAMACHA, pointage des bovins,
- Suivi du bien-être animal de l'animal (enregistrement papier à partir d'une grille d'observation journalière ou spécifique en fonction de l'étude),
- Suivi sanitaire des traitements (ordonnance + protocole de soins sur animaux appareillés ou ayant subi une intervention chirurgicale dans le cadre d'un protocole).

Les mesures instrumentées sont obtenues par l'utilisation d'un ou de plusieurs outils de phénotypage.

Méthode de collecte de données	Données obtenues
Balance automatisée + cage de pesée	Poids
Ruban barymétrique	Estimation du poids
Echographie	Epaisseur de lard, diagnostic de gestation
Distributeurs automatiques de concentrés	Mesure de l'ingestion individuelle
Enregistrement par corpus audiovisuel	Images de comportement, de posture

Les données transformées sont issues de données relevées par observation ou de manière instrumentée qu'il a fallu regrouper et analyser avant leur diffusion.

Prélèvements	Données obtenues
Prélèvements sanguins	Formules sanguines
Prélèvements de fèces	OPG (œufs par grammes), parasitologie
Prélèvements d'urine	Composition chimique (azote)
Prélèvements de jus de rumen	Composition chimique

## Type de données

- Audiovisuelles
- Dataset
- Image

Les données de type dataset sont prépondérantes.

**Observations** : NEC, pointage, suivi quotidien du BEA, comportement.

**Mesures instrumentées** : pesées, échographies, colliers accéléromètres, ingestion, comportement (vidéo).

**Mesures transformées** : bilan alimentaire, analyses de formules sanguines, coproscopies, analyse de jus de rumen.

### *NATURE DES DONNÉES*

**Données de mensurations des animaux vivants** : poids à tout âge ou stade physiologique (pesée ponctuelle des animaux enregistrée en gramme), Note d'état corporel (NEC) : notation de 0 (= animal extrêmement émacié) à 5 (= animal très gras), autopesée : pesée, enregistrée en kilogrammes, à chaque passage volontaire des animaux sur la balance du Distributeur automatique de concentré (DAC).

**Données physiologiques** : suivi du bien-être animal (notation d'absence ou présence d'anomalie pour chaque critère d'une grille d'évaluation en fonction de l'expérimentation), échographie épaisseur de lard, bilan métabolique par des cages à bilan (pesée des intrants et refus, pesée des excréta (urine et fèces).

**Données zootechniques** : généalogie, taille et poids des portées naissance et sevrage, œstrus et insémination.

**Données de parasitologie** : coproscopies, formules sanguines

**Données sanitaires** : ensemble des ordonnances reçues, traitements réalisés ou événements sanitaires réalisés

**Données d'élevage** : croissance des animaux, mesure d'ingestion, de proposé et de refusé en individuel ou en lot en automatique ou pesée manuelle selon besoin et dispositif utilisé.

**Données qualité des produits** : poids des carcasses et pièces de viande, teneur en gras

**Données vidéos de test de comportement** : vidéos, données d'accélération

### *FORMAT DES DONNÉES*

Les données sont enregistrées sous documents papier, sous fichiers tabulés au format xlsx ou csv, pdf, ou sous SICPA espèces ou sanitaire. Les données image ou vidéo sont sous les formats jpg, jpeg, png, H264.

- Documents papier : carnets de mise-bas, suivi de bien-être animal
- Fichier xlsx ou csv → Notation BEA, données d'alimentation/ d'ingestion individualisé, données statistiques, données d'accélération
- Fichier pdf → coproscopies, fèces
- Données sous SICPA sont enregistrées sous format MySQL, PostgreSQL, InfluxDB qui peuvent être exportées sous forme de fichiers ouverts au format texte, à la demande ( ex : poids, NEC, échographie, données alimentation en suivi quotidien, données sanitaires). cf PGD du CATI SICPA
- Format jpg, jpeg, png → photos
- Format H264 → vidéos

## **\*\*formats utilisés\*\***

Format tabulaire : .csv; .xls; .xlsx

Format texte : .txt; .doc; .docx; .pdf

Format photo : .jpg; .jpeg

Format vidéo : H264

## *PÉRIMÈTRE THÉMATIQUE DES DONNÉES*

- Élevage et produits animaux
- Santé et pathologie animales
- Systèmes et pratiques agricoles

## **ENSEMBLE DE DONNÉES LIÉES A L'ENVIRONNEMENT (BIOMASSES, COMPOSITION D'ALIMENTS, TEMPÉRATURE, HYGROMÉTRIE...)**

### *MODE D'OBTENTION DES DONNÉES*

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

Le deuxième grand type de données de recherche collectées par PTEA sont les données dites d'environnement (biomasses, composition d'aliments, températures et hygrométrie).

### **Les données générées par la structure :**

Il s'agit de :

- Données issues des observations visuelles des végétations/cultures et enregistrement des rendements,
- Données issues de mesures instrumentées (station météo),
- Données d'ambiance (SICPA Environnement) : température, humidité, collectées par des systèmes automatisés en temps réel, basés sur un réseau LoRa. Elles arrivent en temps réels, dans un stockage orienté "séries temporelles" dans le datacenter du centre de recherche de Toulouse.

Ces données sont collectées via des saisies papiers, des saisies sur tableurs ou l'utilisation de SICPA Alimentation. Les données sont stockées sous différents formats : papiers, tableurs (.xlsx ou .csv) ou format texte (.txt; .docx; .pdf).

### **Données produites par des tiers :**

Il s'agit essentiellement de :

- Données sur l'analyse de la composition chimique des aliments produits sur l'unité expérimentale

## Origine

- Expérimentation
- Observation

Les observations sont obtenues par relevés visuels des végétations pour le suivi de l'évolution des parcours ou du choix de la parcelle de fauche (ex : herbomètre pour mesurer la hauteur de l'herbe).

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

- Station météorologique : sondes... → données météorologiques
- Capteurs thermiques ou d'humidité → mesures d'ambiances

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 (rendement, hauteur d'herbe)

Enregistrements instrumentés → sondes de température et d'humidité via la station météo du domaine, données d'ambiance via des capteurs en bâtiment (système LORAWAN)

Analyses de laboratoire → composition chimique

### *NATURE DES DONNÉES*

**Données météorologiques** : date et heure de la mesure ; température en degré Celsius (T) ; humidité instantanée en % ; 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 végétation** : relevés de végétation à l'entrée et sortie de parcelles, caractérisation des espèces dominantes, suivi du nombre de jours de pâturage par parcelle et du chargement par parcelle par jour ; relevés de biomasse en kilogrammes de matière sèche par hectare, hauteur d'herbe.

**Données de composition chimique (analyse de laboratoire)** : échantillons fourrage ou d'aliments : % de matière sèche ; MM + MAT + CB + NDF + ADF + ADL en grammes par kilogrammes de matière sèche.

### *FORMAT DES DONNÉES*

- Données météorologiques : fichier csv ou xlsx,
- Données végétation, données d'ambiance, résultats d'analyses de composition chimique : fichier xlsx ou csv.
-

*PÉRIMETRE THÉMATIQUE DES DONNÉES*

- Élevage et produits animaux
- Santé et pathologie animales
- Systèmes et pratiques agricoles

## **DROITS DE PROPRIÉTÉ INTELLECTUELLE**

*QUI DÉTIENDRA LES DROITS SUR LES DONNÉES ET LES AUTRES INFORMATIONS CRÉÉES ?*

INRAE possède les droits, à l'exception d'accord de consortium ou de contrat avec des partenaires privés. Dans le cadre d'un projet, la propriété des données est partagée entre INRAE et les acteurs membres du projet. Les modalités de partage sont définies dans le PGD du projet.

## SENSIBILITÉ DES DONNÉES

### Identification du niveau de sensibilité des jeux de données

Pas de données sensibles.

### 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 la politique décrite dans 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'unité et aux partenaires scientifiques identifiés (PGD CATI SICPA), via par exemple des accès filtrés après authentification et selon le rôle de chacun dans les Systèmes d'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 données que nous produisons ne sont pas directement ou indirectement identifiantes pour des personnes. Si cela devait être le cas, le nom des agents serait anonymisé lors de l'export des données. Pour les données caméras, les agents sont d'abord avertis lorsqu'une expérimentation utilisant une caméra est en cours et leur consentement est recueilli. Aucune image comportant des agents n'est diffusée dans les jeux de données. En revanche, ces images sont tout de même analysées par l'algorithme, qui ne considère que l'animal. Lorsque les images sont diffusées, les agents présents sont floutés de manière à ne pas pouvoir les identifier. Nous nous basons en particulier sur les recommandations de la fiche référentiel vidéosurveillance animaux du projet de recherche des délégués informatique et libertés d'INRAE.

## **PARTAGE DES DONNÉES**

*Y A T'IL UNE OBLIGATION DE PARTAGE (OU A L'INVERSE UNE INTERDICTION OU UNE RESTRICTION) ?*

L'Unité Expérimentale PTEA 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, personnelles ou sensibles).

*QUELLES SONT LES REUTILISATIONS POTENTIELLES DE CES DONNÉES ?*

Les réutilisations des données sont variées.

Elles peuvent servir à :

- Des publications, des valorisations scientifiques et techniques
- Des travaux de recherche et développement (modélisation de systèmes, mise au point d'algorithmes de phénotypage)
- De l'enseignement

*LA LECTURE DES DONNÉES NÉCESSITE-T-ELLE LE RECOURS A UN LOGICIEL OU UN OUTIL SPECIFIQUE ? SI OUI, LEQUEL ?*

Les versions papier peuvent être consultées sur site par tous les agents de l'UE.

Les données d'élevage SICPA sont lues avec le logiciel de développement du même nom (Voir CATI SCIPA) et exportables dans un format ouvert (ex texte).

Les outils bureautiques classiques suffisent pour lire la majorité des données qui sont générées sous fichier de format tabulés ou csv.

Les supports audiovisuels nécessitent des logiciels de visualisation vidéo (type mp4).

Les données statistiques sont lisibles avec des logiciels spécifiques tels que R ou SAS.

### **Comment les données seront-elles partagées ?**

Dans le cadre des projets, les données sont envoyées au partenaire par e-mail ou espace partagé en ligne (Sharepoint, Nextcloud...) ou en local.

Pour les données hors projet, elles sont partagées sur demande.

Pour les données enregistrées sur le SICPA, elles seront partagées selon les droits de consultation du CATI SICPA.

### *AVEC QUI ?*

- Partenaire(s) identifié(s)

Les données associées à des projets peuvent être partagées aux agents INRAE authentifiés (LDAP) et aux partenaires identifiés suivant les projets de recherche.

Les données hors projet et non sensibles peuvent être partagées sur demande.

### *SOUS QUELLE LICENCE ?*

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

## ORGANISATION ET DOCUMENTATION DES DONNÉES

*QUELS MÉTHODES ET OUTILS SONT UTILISÉS POUR ACQUÉRIR ET TRAITER LES DONNÉES, DEPUIS LEUR ACQUISITION JUSQU'A LEUR MISE A DISPOSITION, LEUR ARCHIVAGE OU LEUR DESTRUCTION? UTILISER EVENTUELLEMENT UN LIEN VERS UN SCHEMA ILLUSTRANT LES PROCESSUS*

1. Collectées à l'aide :
  - Papier - Crayon
  - PDA
  - Appareil photo, caméras, drones
  - Stations météorologiques automatiques
  - Sondes
2. Traitées sur :
  - Logiciel, tableur
  - SICPA (espèces, alimentation et sanitaire, environnement...)
3. Stockées sur :
  - Classeur (données format papier)
  - Stockage local individuel (clé USB, disque dur PC)
  - Sharepoint, Nextcloud, espaces partagés en local
  - Serveur INRAE PTEA puis sauvegarde quotidienne Datacenter Toulouse

Des données sont générées par des appareillages automatisés (Distributeurs automatiques de concentrés...), qui disposent d'un logiciel propre. Ces données sont ensuite extraites manuellement vers un fichier de format tabulé.

Étape	Méthode	Outils
Préparation	Rédaction de protocoles	Traitement de texte Word
Acquisition	Procédures, protocoles, modes opératoires	Saisie papier Outils de mesures
Traitement des données	Pré-traitement des données, Vérification, Tri des données aberrantes	PGD SICPA, Logiciels spécifiques aux équipements Tableurs (Excel)
Mise à disposition	Envoi par mail Mise en ligne sur l'espace de partage Saisie sur SICPA	PGD SICPA Mails Espaces partagés
Archivage	Non défini aujourd'hui	
Destruction	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,

AHOL = caractéristiques liées à la santé animale,

EOL = caractéristiques liées à l'environnement,

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

L'ensemble des données qui ont vocation à intégrer les SICPA sont décrites conformément au standard choisi par le CATI (se référer au PGD CATI SICPA).

Pour les autres données de mesure physique, les unités utilisées correspondent aux unités utilisées dans le système international d'unités.

Pour les observations, des grilles d'observation ont été définies en interne (suivi de paramètres comme alimentation et abreuvement des animaux, environnement, santé et comportement des animaux). Les grilles sont accompagnées d'un lexique.

### *UNE DOCUMENTATION COMPLÉMENTAIRE AUX METADONNÉES EST-ELLE NÉCESSAIRE POUR DÉCRIRE LES DONNÉES ET ASSURER LEUR RÉUTILISABILITÉ SUR LE LONG TERME ?*

Pour les données associées à des projets, les protocoles expérimentaux fournissent les métadonnées.

Pour les données hors projet, des fichiers "lisez-moi" sont envisagés

Le Cati SICPA travaille sur l'enrichissement de son catalogue de données, par des liens systématiques des données qu'il gère vers les ontologies ATOL, AHOL, EOL et/ou vers le thésaurus INRAE.

Dans certains cas, des data papers sont rédigés pour accompagner les jeux de données.

### *COMMENT LES FICHIERS DE DONNÉES SONT-ILS GÉRÉS ET ORGANISÉS : CONTRÔLE DES VERSIONS, CONVENTIONS DE NOMMAGE DES FICHIERS, ORGANISATION DES FICHIERS*

Pour les données saisies sous les SI, les informations sont fournies dans le PGD du CATI SICPA.

Pour les autres données, elles sont classées dans des fichiers informatisés par année d'observation et puis par programme de recherche.

Le contrôle des versions d'enregistrement est réalisé par rapport à la date et l'heure d'enregistrement sur le serveur. Pour certains fichiers, les versions sont notées dans le titre du document.

Une réflexion est en cours sur l'organisation des fichiers des données, l'amélioration de la traçabilité et la sécurité des fichiers. Nous réfléchissons à l'adoption d'une convention de nommage simple et harmonisée, qui permettrait également le versioning de type : AAAMMDD\_Nom du projet\_Vx

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

### **Pour la collecte des données :**

- Données instrumentées :

En cas d'acquisition d'un nouvel outil, les agents sont formés à l'utilisation par le fournisseur. Les nouveaux agents recrutés sont formés en interne grâce à l'expertise des anciens.

Le contrôle qualité des données est effectué grâce au système de métrologie de notre démarche qualité.

1. Calibration et étalonnage des outils de mesure. Les balances sont contrôlées par une entreprise externe une fois par an.
2. Acquisition de la donnée suivant une procédure et vérification technique par le collecteur
3. Validation scientifique des données par le demandeur (chercheur)
4. Vérification par le SI au moment de la collecte par un contrôle de vraisemblance, possiblement bloquant (cf PGD CATI SICPA).

- Données d'observation (NEC, FAMACHA : méthode de diagnostic visuel de l'état d'anémie en observant la conjonctive de l'oeil des ovins/caprins) :

Les agents sont formés en interne et une calibration de la façon de noter par notation multiple par plusieurs agents est mise en place.

**Pour la saisie manuelle des données :**

Nous avons mis en place un système de vérification de la transcription des données par double lecture.

Les documents papier sont conservés pour vérification ultérieurement si besoin (pendant 10 ans).

## 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

Nous n'avons pas de système d'information pour notre structure, et le stockage temporaire des fichiers de données n'a pas fait l'objet d'analyse de risque. Ces fichiers sont stockés sur serveur ou ordinateur professionnel. Ils dépendent de l'analyse de risque du SI du Centre Antilles Guyane.

Pour le CATI SICPA, certains systèmes d'information sont en cours de procédure d'homologation (SICPA Sanitaire & Alimentation, SICPA Expérimentation).

La politique de sécurité des systèmes d'informations d'INRAE (PSSI INRAE) sera réalisée ultérieurement.

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

Les données sont stockées sur :

- Les cahiers de laboratoire (stockage papier)
- Le CATI SICPA,
- Le réseau partagé de l'UE PTEA,
- Ponctuellement sur disques durs ou clés USB.

Certaines données sont stockées de façon temporaire sur les outils de collecte :

- DAC GENSTAR pour la consommation alimentaire (animal, durée repas, quantité, poids, détails visites cumul) : automate/sauvegarde Local PC/Dossier partage lecteur réseau.
- GREEN FEED pour l'analyse de la composition des gaz érucés par les ruminants (CH<sub>4</sub>, O<sub>2</sub>, CO<sub>2</sub>, Température, air Flow) : sauvegarde local/ base de données greenfeed.c-lockinc.com.
- GREEN FEED pour la consommation alimentaire (animal, rationnement, consommation totale, paramétrage courbe) : automate/sauvegarde Local PC/Dossier partage lecteur réseau.

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

Nos productions principales concernent des données non sensibles et non chiffrées. Pour les données à caractère personnel, des précautions particulières sont prises (floutage préalable au transfert...).

## Quelle est la volumétrie actuelle et prévisionnelle ?

Stockage :

- Sur les bases du CATI SICPA : nos données sont intégrées dans les bases du CATI qui sont estimées pour l'ensemble des unités expérimentales animales à 60 Mo pour le SICPA Espèces et 10 Go pour le SICPA Sanitaire,
- Sur le serveur de stockage en réseau (espace partagé ou NAS) pour UE PTEA et UR ASSET : 157 Go occupés et 570 Go libres,
- Sur les serveurs gérés par la DSI, chaque agent dispose d'un espace dédié de 15 Go pour la sauvegarde des données individuelles. Le stockage actuel sur les serveurs de la DSI est de 1Go, l'espace libre est de 200 Go. La croissance est estimée à 100 Mo par an.

**\*\*Acronyme\*\***

NAS → Network attached storage (serveur de stockage en réseau)

## 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 ?

Les modalités de la charte des infrastructures de recherche à INRAE sont appliquées à l'unité.

La politique mise en place à PTEA est d'enregistrer sur le serveur de stockage en réseau (NAS) qui n'est accessible qu'en déclarant les appareils dédiés (pas d'ordinateur personnels) ou en utilisant le VPN mis à disposition par le service informatique et par l'identification des agents (ldap). Les données du NAS sont sauvegardées tous les jours et ces sauvegardes sont externalisées sur un site distant. Cela permettrait en cas de grosse défaillance de pouvoir les reconstituer dans le pire des cas à J-1.

L'entité hébergeant physiquement les données applique la PSSI de l'établissement. Les différentes mesures applicables à l'unité sont en cours de mise en œuvre.

Elle n'a pas de plan d'assurance 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 ?

Les données sont partagées avec des partenaires identifiés participant au projet : des échanges ont lieu entre les partenaires et l'UE PTEA lors de l'élaboration de chaque projet de recherche. Les discussions portent notamment sur la sécurisation, la conservation, l'accès, la diffusion ou l'exploitation des données en fonction des contraintes scientifiques et juridiques. Les modalités retenues sont ensuite mentionnées dans le PGD du projet. Dans la plupart des cas, une copie des données est transférée aux partenaires chercheurs, qui assurent les droits de diffusion et d'exploitation des données jusqu'à la publication.

Les autres personnes hors CATI et hors projet doivent faire une demande de collaboration spécifique.

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

Sécurité – intégrité des données

Les événements liés à la production de données qui pourraient faire varier leur valeur ou leur interprétation sont notés dans les cahiers de laboratoire.

Les bases de données du CATI SICPA sont sauvegardées tous les jours et ces sauvegardes sont externalisées dans un coffre. Cela permettrait en cas de grosse défaillance de pouvoir les reconstituer dans le pire des cas à J-1. Un PRA géré par le CTIG (Centre de Traitement de l'Information Génétique animale), testé régulièrement (plan de reprise d'activité), inclut ces systèmes d'informations.

**En cas d'incident**, les données peuvent être récupérées partiellement ou totalement en fonction de la gravité de l'incident, notamment :

- Fichier corrompu : réparation du fichier ; récupération des données même partielle ; réutilisation des versions antérieures stockées dans le dossier archives de l'expérimentation correspondante.
- Suppression accidentelle de fichier : vérification des versions antérieures archivées ; vérification des supports temporaires de stockage (clé USB, ordinateur de terrain) ; utilisation du service informatique du centre pour réexporter le fichier et/ou dossier en question malgré la perte d'informations saisies le jour même. Il faut cependant que la suppression soit antérieure à 30 jours.

Une fois validée, les données de production sont accessibles en lecture seule aux personnes autorisées.

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

- Non

## ARCHIVAGE ET CONSERVATION DES DONNÉES

*QUELLES SONT LES DONNÉES A CONSERVER SUR LE MOYEN OU LE LONG TERME ET QUELLES SONT LES DONNÉES A DÉTRUIRE ?*

Les données stockées dans le CATI sont conservées à long terme notamment les données généalogiques. Pour les données hors CATI, les discussions ont lieu au cas par cas pour identifier les données patrimoniales à conserver à long terme et les données expérimentales qui seront transmises au chercheur et conservées 10 ans à l'unité expérimentale.

*SUR QUELLE PLATEFORME D'ARCHIVAGE PÉRENNE SERONT ARCHIVÉES LES DONNÉES A CONSERVER SUR LE LONG TERME ? SINON, QUELLES PROCÉDURES SERONT MISES EN PLACE POUR LA CONSERVATION A LONG TERME ?*

Les données à long terme seront conservées sur le CATI (Cf PGD du CATI)

*QUELLE EST LA DURÉE DE CONSERVATION DES DONNÉES ?*

(Cf PGD du CATI)

*QUELLES GARANTIES DE FINANCEMENT COUVRIRONT LES COÛTS ASSOCIÉS A LA CONSERVATION A LONG TERME ?*

Les garanties de financement pour couvrir les coûts associés à la conservation à long terme sont les modalités de financement du CATI (Cf PGD du CATI). Elles sont garanties par le CTIG et ses départements de tutelle GA et PHASE.





**Centre Antilles Guyane**

Domaine Duclos  
97170 Petit-Bourg  
(0)5 90 25 59 00

Rejoignez-nous sur :



<https://www.inrae.fr/centres/antilles-guyane/>

**Institut national de recherche pour  
l'agriculture, l'alimentation et l'environnement**



**RÉPUBLIQUE  
FRANÇAISE**

*Liberté  
Égalité  
Fraternité*

**INRAE**