
DMP du projet "Unité Expérimentale PEIMA"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "INRAE - Trame Structure" 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 "Unité Expérimentale PEIMA"
Livrable	v0
Version	Version initiale
Domaines de recherche (selon classification de l'OCDE)	Agriculture, forestry, and fisheries, Animal and dairy science
Langue	fra
Date de création	2020-04-23
Date de dernière modification	2024-01-26
Identifiant	10.15454/1.5572329612068406E12
Type d'identifiant	DOI
Licence	Creative Commons Zero v1.0 Universal
Plans de gestion en lien avec le projet	<ul style="list-style-type: none">PGD-SICPA : https://doi.org/10.57745/ACTK8H

Renseignements sur le projet

Titre du projet	Unité Expérimentale PEIMA
Acronyme	DMP - UE PEIMA

Résumé

La Pisciculture Expérimentale INRAE des Monts d'Arrées (PEIMA) intervient dans des programmes nationaux de recherche pluridisciplinaire qui abordent l'ensemble des composants de la filière salmonicole d'eau douce, depuis l'oeuf jusqu'au produit de consommation.

Les expérimentations mises en place en biologie des poissons participent à une meilleure compréhension de :

- la physiologie (physiologie de la reproduction, de la croissance, analyses du comportement, adaptation à des milieux changeants, etc.)
- la génétique (structure du génome, déterminisme génétique, sélection, etc.)
- la nutrition (besoins nutritionnels, épargne des matières premières d'origine marine, etc.)
- la défense contre les maladies (virologie moléculaire, amélioration des systèmes de défense spécifiques et non spécifiques).

La PEIMA offre également une collection de salmonidés d'élevage unique au monde grâce à la sélection et au maintien de lignées aux spécificités originales (lignées isogéniques homozygotes). Ces souches et lignées seront mises à disposition de la Cryobanque nationale, dont l'INRAE est partenaire, à travers le projet [CRB-Anim](#).

Sources de financement

- INRAE :

Partenaires

- Laboratoire de Physiologie et Génomique des Poissons (199717864K)
- Génétique Animale et Biologie Intégrative (200917790C)
- Nutrition, Métabolisme, Aquaculture (199917825J)
- Plateforme d'Infectiologie Expérimentale des Rongeurs et Poissons (198917775Y)

Produits de recherche :

1. Expérimentation animale (Jeu de données)
2. Expérimentation système (Jeu de données)

Contributeurs

Nom	Affiliation	Rôles
KERNEIS Thierry	INRAE	<ul style="list-style-type: none"> • Responsable du plan de gestion de données
LARRANAGA Nicolas	INRAE	<ul style="list-style-type: none"> • Coordinateur du projet
Laurent LABBE		<ul style="list-style-type: none"> • Personne contact pour les données (Exp animale, Exp système)

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.

DMP du projet "Unité Expérimentale PEIMA"

Informations sur la structure

Nom de la structure

UE 0937 PEIMA Unité Expérimentale sur les Systèmes d'Elevage Aquacoles
Pisciculture Expérimentale INRAE des Monts d'Arrées

Type de structure

- ISC (Infrastructure Scientifique Collective)
- Unité de recherche, Unité ou Installation Expérimentale
- ISC (Infrastructure Scientifique Collective)
- Unité de recherche, Unité ou Installation Expérimentale

<https://www6.rennes.inrae.fr/peima>

Identifiant de la structure

Préciser le fournisseur de l'identifiant (ISNI, VIAF, FundRef, DataCite...).

10.15454/1.5572329612068406E12 (DataCite) 199217743A (RNSR)

Responsabilités dans la structure

Nom, Prénom	Courriel	Rôle
LABBE Laurent	laurent.labbe@inrae.fr	Directeur d'unité
GOARDON Lionel	lionel.goardon@inrae.fr	Responsable infrastructure
KERNEIS Thierry	thierry.kerneis@inrae.fr	Responsable technique
TERRENNE Tina	tina.terrenne@inrae.fr	Référent données opérationnel (RDO)

Etablissement(s) tutelle(s)

INRAE - Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement

Département de rattachement INRAE (ou anciens départements Inra)

- GA
- PHASE
- GA

- PHASE

Le département PHASE étant le département pilote.

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

INRAE

Informations sur le plan de gestion

DOI (*version publiée du plan de gestion*)

à prévoir

Historique des versions

Date	n° de version	Status	Auteur	Affiliation de l'auteur (se reporter à l'annuaire INRAE)	Validé par	Validé le
23/04/2020	0.1		Thierry Kerneis & Tina Terrenne	UE 0937 PEIMA Unité Expérimentale sur les Systèmes d'Elevage Aquacoles barrage du drennec 29450 SIZUN		

Présentation générale des données

Expérimentation animale

Mode d'obtention des données

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

Les données générées par la structure sont issues des observations visuelles et de phénotypages instrumentés ou non. Ces données sont collectées via tableur avec ou sans macro ou saisie papier. Les données sont ensuite retranscrites sur tableur, fichier texte ou sur les logiciels d'élevage Ecu2000 et SI SICPA.

Les données générées par un tiers sont conservées en version papier, informatiques et retranscrites sur tableur.

Origine

- Expérimentation
- Analyse
- Simulation, modélisation
- Expérimentation
- Analyse
- Simulation, modélisation

Le jeu de données "expérimentation animale" est produit en accompagnement des données produites par les projets concernant l'expérimentation animale (demandeurs externes à l'unité : chercheurs, etc.), ce sont les données concernant les poissons et leur environnement direct et indirect.

Type de données

- Dataset
- Audiovisuel
- Image
- Text
- Dataset
- Audiovisuel
- Image
- Text

Prépondérance des données de type dataset et text, plus rarement image et audiovisuel.

Nature des données

Environnement :

Données générées. Obtenues par analyse et expérimentation. Type de données : dataset. Périmètre thématique : métadonnées, données d'élevage.

Température, nitrite, nitrate, ammonium, O2, saturation O2, débit global d'eau, débit d'eau du bassin, niveaux d'eau, profil thermique du lac, cyanobactéries.

Biométrie du poisson vivant :

Données générées, données produites par un tiers (achat de données). Obtenues par analyse et expérimentation. Type de données : dataset, texte, image. Périmètre thématique : données d'expérimentation.

Poids, longueur, taux de gras, épaisseur, stade physiologique.

Biométrie du poisson abattu :

Données générées. Obtenues par analyse et expérimentation. Type de données : dataset, texte, image. Périmètre thématique : données d'expérimentation.

Poids (total, viscères, carcasse, foie, gonade, filet, tête, œufs restant dans la cavité abdominale), longueur, taux de gras, colorimétrie, pH, conformation (calcul), épaisseur, matière sèche.

Zootechne :

Données générées. Obtenues par analyse et expérimentation. Type de données : dataset, texte. Périmètre thématique : données d'élevage.

Aliment distribué, type d'aliment, granulométrie, aliment refusé, taux de rationnement, croissance prévisionnelle, croissance effective, densité, mortalité, entrées et sorties d'animaux, traitements externes, traitements vétérinaires (temps, molécule, quantité), traitements hormonaux (temps, molécule, quantité), comportement, mouvements, état sanitaire.

Comportement :

Données générées. Obtenues par analyse et expérimentation. Type de données : dataset, audiovisuel, texte. Périmètre thématique : données d'expérimentation.

Bien-être animal, prise alimentaire, état général.

Reproduction :

Données générées, données produites par un tiers (achat de données). Obtenues par analyse et expérimentation. Type de données : dataset, texte, image. Périmètre thématique : données d'expérimentation.

Poids (ponte, liquide coelomique, 100 œufs), sexe, état de maturation sexuelle, diamètre des œufs, taux de fécondation dont nombre (œufs, blancs, non fécondés, oeuvillés, haploïdes, éclos, non éclos), taux de motilité du sperme, temps d'activation du sperme, volume de sperme.

Métadonnées :

Données générées. Obtenues par analyse et expérimentation. Type de données : dataset. Périmètre thématique : métadonnées.
Propriété du document (date de création, auteur, emplacement du fichier), version (i.e. v1.1), date et lieu de collecte (issu de fichier avec macro ou en saisie manuelle), protocole de collecte, appareil de mesure, (i.e. balance LE10001), informations de qualité (étalonnage et conformité) sous forme d'enregistrements de vérification avant mesure ou étalonnage, fiche de vie du matériel, généalogie du matériel biologique utilisé.

Format des données

Format tabulaire : .csv; .xls; .xlsx

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

Format photo : .jpg; .jpeg

Format vidéo : .mp4; .avi

Données d'élevage : Ecu2000 et SICPA (.fic, .ndx, .csv)

Documents papier

Périmètre thématique des données

- Animal Breeding and Animal Products
- Fishes and Aquaculture
- Food and food processing
- Animal Health and Pathology
- Animal Breeding and Animal Products
- Fishes and Aquaculture
- Food and food processing
- Animal Health and Pathology

Majoritairement "Fishes and Aquaculture". Les autres thématiques sont également abordées par les jeux de données obtenus par la structure.

Expérimentation système**Mode d'obtention des données**

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

Les deux types d'obtention de données existent au sein de l'unité avec une grande majorité des données générées par la structure. Certaines données spécifiques (dosage de micro-nutriments, population bactérienne) sont obtenues avec l'utilisation de services et structures tiers.

Origine

- Analyse
- Expérimentation
- Simulation, modélisation

- Analyse
- Expérimentation
- Simulation, modélisation

Trois origines déterminées : analyse, expérimentation et modélisation.

Le jeu de données "expérimentation système" est produit par les expérimentations menées en interne (projets menés par la structure hors demandes externes). Les données sont traitées par les intervenants internes à l'unité.

Type de données

- Audiovisuel
- Image
- Text
- Dataset
- Audiovisuel
- Image
- Text
- Dataset

Prépondérance des données de type dataset et text, plus rarement image et audiovisuel

Nature des données

Environnement :

Données générées. Obtenues par analyse et expérimentation. Type de données : dataset. Périmètre thématique : métadonnées, données d'élevage.

Il s'agit d'analyses d'eau réalisées à fréquence régulière en différents points de la pisciculture dont l'objectif est de caractériser la physico-chimie de l'eau et de s'assurer de bonnes conditions d'élevage.

L'acquisition des données se fait en deux étapes. Une première dans les élevages, correspondant aux prélèvements et l'autre au laboratoire.

Paramètres analysés : nitrites, nitrates, ammonium, oxygène dissous (et taux de saturation), fer, turbidité, phosphore dont orthophosphates, conductivité, dioxyde de carbone dissous, DBO et DCO, matières en suspension, pH.

Les températures sont relevées deux fois par jour à divers endroits de la pisciculture et relevées en continu dans plusieurs infrastructures expérimentales.

Biométrie :

Données générées, données produites par un tiers (achat de données). Obtenues par analyse et expérimentation. Type de données : dataset, texte, image. Périmètre thématique : données d'expérimentation.

Poisson : poids, longueur, population bactérienne (ext.).

Végétal : poids (feuille, racine, total), matières sèches, microéléments (externe), population bactérienne (ext.), gain de biomasse, croissance (foliaire, racinaire).

Zootecnie :

Données générées. Obtenues par analyse et expérimentation. Type de données : dataset, texte. Périmètre thématique : données d'élevage.

Aliment distribué, type d'aliment, granulométrie, aliment refusé, taux de rationnement, croissance prévisionnelle, croissance effective, densité, mortalité, entrées et sorties d'animaux, traitements externes, traitements vétérinaires (temps, molécule, quantité), traitements hormonaux (temps, molécule, quantité), comportement, mouvements, état sanitaire.

Métadonnées :

Données générées. Obtenues par analyse et expérimentation. Type de données : dataset. Périmètre thématique : métadonnées.

Propriété du document (date de création, auteur, emplacement du fichier), version (i.e. v1.1), date et lieu de collecte (issu de fichier avec macro ou en saisie manuelle), protocole de collecte, appareil de mesure, (i.e. balance LE10001), informations de qualité (étalonnage et conformité) sous forme d'enregistrements de vérification avant mesure ou étalonnage, fiche de vie du matériel, généalogie du matériel biologique utilisé.

Format des données

Format tabulaire : .csv; .xls; .xlsx
Format texte : .txt; .doc; .docx; .pdf
Format photo : .jpg; .jpeg
Format vidéo : .mp4; .avi
Données d'élevage : Ecu2000 et SICPA (.fic, .ndx, .csv)
Documents papier

Périmètre thématique des données

- Animal Breeding and Animal Products
- Animal Health and Pathology
- Fishes and Aquaculture
- Food and food processing
- Animal Breeding and Animal Products
- Animal Health and Pathology
- Fishes and Aquaculture
- Food and food processing

Majoritairement "Fishes and Aquaculture". Les autres thématiques sont également abordées par les jeux de données obtenus par la structure.

Droits de propriété intellectuelle

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

La PEIMA possède les droits de propriété intellectuelle des données produites hors projets.

Dans le cadre des projets, la propriété des données est partagée entre la PEIMA et les acteurs membres du projet. C'est le PGD du projet qui la détermine.

Confidentialité

Identification des jeux de données contenant des données confidentielles

Pour le moment, nous n'identifions pas de jeux de données contenant des données confidentielles (hors projets).

Quelles sont les mesures prises et les normes auxquelles il est nécessaire de se conformer pour garantir cette confidentialité ?

Dans le cadre de la confidentialité liée à un projet (partenaire privé par exemple), elle 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'unité.

A l'avenir des questions sur l'accessibilité devront être posées en fonction du jeu de données concerné.

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 ?

Les données qui seront mises à disposition ne seront pas concernées par le RGPD. Elles ne seront par directement ou indirectement identifiantes pour les personnes.

Les données collectées et mises à disposition pour une analyse de second niveau le seront uniquement après l'utilisation de celles-ci par les acteurs de projets auxquelles elles peuvent être associées.

Partage des données

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

Il n'y a pas de restriction pour le partage des jeux de données obtenus hors projet.

La structure INRAE met en place une politique de partage des données : le plus ouvert possible, le plus fermé que nécessaire.

La politique institutionnelle précise qu'une donnée obtenue à plus de 50% avec des financements publiques devient un document administratif et doit donc être rendue publique. Cela concerne les données finalisées uniquement. Cela nous permettra à terme de partager les jeux de données produits hors projets aux environs d'une fois par an et les données pouvant être liées à des projets (données environnementales) pourront être mises à disposition après l'embargo défini dans le projet.

Quelles sont les réutilisations potentielles de ces données ?

Les données collectées pourraient être réutilisées notamment pour :

- Incrémenter des simulateurs d'expérimentations (expérimentation virtuelle) avec les données zootechniques (i.e machine learning).
 - Trouver des corrélations non étudiées avec l'analyse de premier niveau (i.e. type d'aliment et qualité de reproduction).
 - Identifier et quantifier les dérives génétiques, physiologiques et autres des entretiens de collections.
 - Définir un catalogue des caractéristiques physiques, biologiques, génétiques et physiologiques de chacune des collections entretenues.
-

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 ne peut se faire qu'avec le logiciel Ecu2000 et en ayant accès aux informations stockées sur le serveur de la PEIMA (présence sur le site ou avec le VPN fourni par le service informatique).

- La lecture des données d'élevage SICPA peut se faire avec le logiciel en cours de développement du même nom. A terme, les acteurs des projets auront à minima un droit d'accès aux informations zootechniques en temps réel.

- Pour les fichiers numériques : les logiciels de la suite Office sont utilisés à la PEIMA, notamment Excel et Word. Pour la consultation la plupart des logiciels de traitement de texte et tableur permettent la lecture des fichiers. Pour la modification, la suite Office est indispensable. Les représentations cartographiques nécessitent un logiciel comme Xmind pour la lecture.

- Pour les versions papier : la consultation sur site est possible par tous les agents de la PEIMA.

Comment les données seront-elles partagées ?

Publication sur le dataverse Recherche Data Gouv

Les jeux de données brutes seront à terme liés à des métadonnées générées et renseignées autour de ces jeux et seront publiés sur le portail data inrae dans le dataverse Recherche data gouv -> Data INRAE -> UE PEIMA dans les datasets dédiés.

La structure des datasets sera prochainement définie mais probablement de type

- dataverse : UE PEIMA
 - dataset : Expérimentation Animale
 - environnement
 - température eau rivière
 - température eau source
 - physicochimique de l'eau
 - méthodologie technique (MO)
 - animal
 - zootechnie
 - traitement sanitaire
 - alimentation
 - généalogie des collections
 - dataset : Expérimentation Système
 - expérimentation 1
 - environnement
 - végétal
 - animal
 - expérimentation 2
 - environnement
 - végétal
 - animal
 - etc.

Avec qui ?

- Agents Inra authentifiés
- Agents Inra authentifiés

Certains jeux de données seront, dans le futur, diffusés en open acces. Dans un premier temps, l'accès aux jeux de données sera permis aux agents de l'institut identifiés. Afin de prendre en main la diffusion et le partage des jeux de données sur le dataverse recherche data gouv, l'option de demande de consultation (formulaire de contact demandant une autorisation systématique) sera présente sur chaque jeu de donnée versé dans l'entrepôt.

Sous quelle licence ?

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

Le format de diffusion des jeux de données sera probablement un format libre de réutilisation de type CC-BY.

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

De nouvelles données sont collectées en suivant les protocoles mis en place pour chacune d'entre elles.

1- Données d'élevage

Les méthodes de collecte des données d'élevage sont rassemblées dans la *procédure gestion d'élevage*. Les données sont enregistrées via les logiciels de gestion d'élevage Ecu2000 (.fic et .ndx) et prochainement SICPA (.csv). Une saisie papier est également effectuée par le responsable du lot de poissons. Il y a également la possibilité d'extraire ces données en fichier texte (.txt) ou en tableur (.xls).

2- Données expérimentales

Les modes de collecte de données d'expérimentation suivent les protocoles disponibles dans le dossier informatique "assurance qualité recherche". Généralement ces données sont rassemblées dans des cahiers de laboratoire et sur tableur (.xls et .xlsx) ou fichier texte (.txt, .doc et .docx) puis sauvegardées sur le serveur.

Dans le détail nous retrouvons les données suivantes :

Température : les températures de l'eau de source et de rivière sont collectées selon la *procédure gardes et astreintes* (deux mesures par jour) à l'aide de thermomètres à alcool. Elles sont saisies sur les logiciels d'élevage (Ecu2000 et SICPA) une fois par semaine. Une version papier est conservée et est consultable par tous les agents de l'unité pour le mois en cours car nécessaire pour adapter les conditions d'élevage. Les archives papier sont conservées sans limite de temps.

Les valeurs mesurées en continu par des sondes température sur le logiciel de supervision concernant les systèmes recirculés (Ar Ster et CR) ne sont pas sauvegardées régulièrement. Elles ne le sont que durant les expérimentations système, où elles sont exportées dans un tableur et sauvegardées sur le serveur. Seules deux valeurs journalières sont conservées et suivent le même traitement que pour l'eau de source et de rivière.

Les températures issues des sondes dans les chambres froides de l'atelier de transformation sont consultables sur le site internet du fournisseur. Elles sont sauvegardées sur le site et régulièrement téléchargées et stockées sur le serveur.

Il est aussi possible d'enregistrer la température en continu pour certaines expérimentations. Celle-ci est alors compilée dans un tableur et sauvegardée sur le serveur.

Analyse d'eau : Les analyses réglementaires (nitrites, nitrates, ammoniac, O₂ dissous et taux de saturation), sont réalisées une fois par semaine en étiage (juillet et août) et une fois par mois le reste de l'année. Elles sont réalisées en trois points dans et hors de la pisciculture et sont stockées dans un tableur annuel. Chaque année le fichier est archivé et un nouveau fichier est ouvert.

Oxygène et taux de saturation : les valeurs d'oxygène dissous sont prélevées une fois par semaine (environ) sur chacun des bassins en fonctionnement. Elles sont conservées en version papier dans le classeur "mesures oxygène" sur site. Une copie est à disposition des agents pour consultation et modification des conditions d'élevage si besoin jusqu'au contrôle suivant.

Il est également possible de collecter des valeurs d'oxygène dissous dans le cadre d'expérimentations. Elles sont alors exportées dans un tableur puis enregistrées dans le serveur sur le dossier correspondant à l'expérimentation. Elles sont également transmises (en physique ou par mail) au chercheur en charge de l'expérimentation.

Il y a également des sondes à oxygène dans les systèmes recirculés, les valeurs mesurées en continu par ces sondes sont affichées en temps réel sur le système de supervision et sont consultables à tout moment. Elles ne sont toutefois exportées et sauvegardées que pour des expérimentations spécifiques. Si nous sommes hors d'une expérimentation, les données sont conservées au niveau de l'ordinateur de supervision pendant 1 mois environ sans toutefois être sécurisées.

L'appareil utilisé pour la mesure est indiqué sur le cahier de laboratoire correspondant à l'expérimentation ou sur la feuille de saisie papier.

Fer : pendant les périodes de l'année concernées par l'apparition de fer dans l'eau de rivière, des analyses de suivi de taux de fer sont réalisées jusqu'à trois fois par semaine. Celles ci sont consignées dans un cahier de laboratoire mis à disposition des agents.

Turbidité, phosphore dont orthophosphates, conductivité, dioxyde de carbone dissous, DBO et DCO, matières en suspension, pH, nitrites, nitrates, ammoniac : ces données sont collectées dans le cadre d'expérimentations. Elles sont alors stockées dans des tableurs (.xls et .xlsx) dans les dossiers des expérimentations correspondantes sur le serveur. Les appareils utilisés sont indiqués si besoin dans les fichiers eux même.

Poids du poisson : collecté sur les lots de poissons (échantillon de population ou poids individuel) toutes les 2 à 6 semaines en suivant la *procédure gestion d'élevage*, l'information est saisie sur Ecu2000 et sur les feuilles de lots papier. Actuellement l'appareil utilisé pour peser les poissons dans le cadre de l'élevage n'est pas indiqué systématiquement.

Poids d'œufs, de liquide cœlomique, de 100 œufs : ces données sont collectées dans le cadre d'expérimentations spécifiques ou pour l'expérimentation permanente entretien de lignées isogéniques. Elles sont sauvegardées sur le serveur et transmises en main propre ou par mail au chercheur concerné. Dans le cadre de l'expérimentation lignées isogéniques ces données sont stockées sur le serveur uniquement.

Poids (poisson, viscère, carcasse, tête, foie, gonade, filet, œufs restants dans la cavité abdominale), longueur, colorimétrie, pH, résistance à la pénétration, conformation (photo), épaisseur sur poisson abattu : mesures prises dans le cadre d'expérimentations. Les appareils utilisés sont généralement précisés sur les fichiers ou les cahiers de laboratoire concernés.

Taux de gras : cette mesure peut-être prise sur poisson abattu ou vivant. Le nom de l'agent réalisant la mesure est noté et stocké en

interne et les valeurs obtenues lors des mesures sont sauvegardées sur le serveur.

Végétaux : poids (feuille, racine, total), matière sèche, microélément (ext.), population bactérienne (ext.), gain de biomasse, croissance (foliaire et racinaire). Ces données prises dans le cadre d'expérimentations système sont conservées en version papier et consignées dans des tableurs.

Entrées et sorties d'animaux : pour chaque entrée et sortie de poissons vivants il y a un suivi sur le logiciel de gestion et sur papier. Pour chaque sortie de poissons abattus (transformation) et morts (équarrissage), suivi papier avec transmission électronique (pour l'équarrissage).

Autres données : propriétés des documents (date de création, auteur, emplacement de fichier, volume, version pour les documents qualité) en saisie semi-automatique dans un fichier à la racine du sous dossier, date et lieu de collecte sur le fichier concerné, protocole de collecte de la donnée utilisé, appareil utilisé lors de la collecte, informations de qualité (étalonnage de l'appareil/conformité). Celles-ci ne sont pas systématiquement reliées aux jeux de données produits.

Étape	Méthode(s)	Outil(s)
Préparation	- Rédaction de protocoles et modes opératoires et mise à disposition à tous les agents de l'unité. - Formation interne aux gestes techniques.	- Traitement texte
Acquisition	- Procédures, protocoles et modes opératoires validés par le responsable qualité. - Pour les échantillons : collecte, conditionnement, identification.	- Saisie papier / ordinateur terrain - Outil de mesure (balance, etc.)
Traitement des données	- envoi aux acteurs des projets (exp. animale) ou analyse de premier niveau à des fins de présentation (exp. système)	- tableur
Mise à disposition	- à venir : publication sur Recherche data gov - actuellement : accessible aux acteurs des projets concernés et aux agents de la PEIMA	- Entrepot de données serveur/mail
Archivage	Dossier "archive" sur le serveur	Serveur
Réutilisation	Prochainement : quand publié sur Recherche data gov	NA
Destruction	Non définie	NA

Acquisition : Les données sont acquises en suivant des procédures, protocoles et modes opératoires validés par le responsable qualité du site qui sont disponibles pour tous les agents travaillant sur l'unité et consultables à tout moment. Les formations en interne permettent d'affiner les compétences et connaissances nécessaires à la bonne réalisation des mesures. Un certain nombre de métadonnées sont collectées en même temps. Une procédure de vérification des appareils de mesure (métrologie) est en cours de construction au sein de l'unité.

Traitement des données : Peu de traitements de données sont actuellement réalisés sur les jeux de données produits hors projet. Ceux qui sont fait correspondent à une analyse de 1er niveau pour communiquer (généralement en interne) sur les résultats obtenus. Ces communications se font sous différentes formes que ce soit poster, résumé, ou communication orale lors de présentations.

Mise à disposition : La mise à disposition des données se fait aux partenaires des projets pour les données qui sont produites dans ce cadre. A cela peut s'ajouter la transmission de données d'élevage générales, d'environnement ou d'autres données comme les protocoles utilisés pour la collecte ou l'état de conformité métrologique des appareils de mesure. Les données sont également disponibles pour les agents INRAE présents et travaillant au sein de l'unité.

Archivage : Deux archivages sont en vigueur sur l'unité à savoir l'archivage papier dans une salle dédiée et l'archivage dans des dossiers informatiques sur le serveur généralement nommés "archives".

Destruction : La politique de l'unité ne définit pas de durée de vie pour la majorité des données produites. Celles-ci sont conservées en format papier ou en version numérique (et souvent les deux). Quelques données ont une durée de vie fixée par la réglementation appliquée en matière d'élevage et de suivi sanitaire des lots.

Les données suivantes : ordonnance vétérinaire, bon de transport et document qualité sont conservées 5 ans avant suppression des fichiers sur le serveur et destruction des versions papier.

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 ?

Métadonnées	Origine, mode de production des métadonnées (ex : saisie manuelle, annotation automatique...)	Standard, Vocabulaires associés	Conditions ou fréquence de la mise à jour (si applicable) (ex : changement de l'accessibilité)
Propriétés des documents (date de création, auteur, emplacement de fichier, volume, version)	semi automatique avec actualisation manuelle		actualisation une fois par an
Date et lieu	manuelle sur le fichier		
Protocole	manuelle		révision des documents une fois par an
Nom de l'appareil	automatique pour les fichiers avec macro, manuel pour les autres		
Etalonnage de l'appareil	manuelle		au minimum une fois par an
Ontologie/thesaurus	manuelle (sur le fichier)	ATOL, EOL, AHOL, thesaurus inrae	avant publication sur data INRAE (quand applicable)
Personne qui fait la mesure	manuelle (sur le fichier)		
Contrôle métrologique	manuelle		
Lot/ individu/ expérimentation concerné.e.s	manuelle		
Type de mesure (individuelle/collective)	manuelle		
Valeurs test	manuelle		
Fichier "lisez moi" à la racine du dataset			

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

Il est préférable que les données soient associées aux ontologies ATOL, AHOL et EOL.

Si la donnée collectée ne correspond pas aux standards de collecte décrits dans ces ontologies, une documentation complémentaire est nécessaire en particulier sur le protocole de collecte. Elle est disponible sur demande par les utilisateurs des données que nous collectons (partenaires des projets). Cette documentation, validée par le responsable qualité de l'unité est stockée sur le serveur.

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

Contrôle des versions

- Pour les documents qualité (MO, ENR, PRO) : selon *PRO gestion documentaire* à savoir un suivi des modifications dans le nom du fichier en première page du document et un archivage des versions précédentes dans le dossier archive de l'assurance qualité recherche.

- Pour les documents d'expérimentation : conservation du document précédent dans le dossier correspondant à l'expérimentation et transmission éventuelle (en physique ou par mail) du document au chercheur en charge de l'expérimentation. Pour les documents comme les bases de données, l'ancienne version est placée dans l'archive du dossier d'expérimentation correspondant.

- Pour les données issues d'élevage : archivage des feuilles de lots papier dans le classeur de l'expérimentation correspondant avec numéro de page et année de début de renseignement de la page sur chaque feuille. Quand le lot est éliminé ou l'expérimentation finie, les feuilles de lots sont archivées dans des boîtes d'archives nommées et conservées sur place.

Convention de nommage

Pour les documents qualité (MO, ENR, PRO) : selon *PRO gestion documentaire* sous la forme :

Type de document (mode opératoire MO, enregistrement ENR, procédure PRO) + intitulé + version (i.e. MO_anesthésie_PEIMA_v1.1).

Pour les documents d'expérimentation :

Nom du projet/acronyme + type de collecte (passage, reproduction, abattage, fumage, pesée, base de données, etc.) + date de collecte

ou dernière modification + informations complémentaires (i.e. QPO_marquage_LGM_22-09-20).
Les fichiers ne sont pas tous nommés de la même manière. Une uniformisation est à prévoir.

Pour les documents d'élevage : numéro de lot en haut de la feuille papier, année de début de saisie et numéro de feuille.

Organisation des fichiers

Sur le serveur les fichiers sont répartis par zone, explicitées sur le fichier *cartographie organisationnelle*.

Ces 7 zones permettent le classement des informations avec notamment : administratif, infrastructures expérimentales, points critiques, expérimentations (avec un dossier pour chaque expérimentation), appui scientifique (avec assurance qualité recherche et métrologie), appui technique (gestion stocks, fiches de vie), vie et travail à la PEIMA (plannings) et utilisateurs. Pour chaque sous dossier d'intérêt, un tableur liste l'ensemble des fichiers disponibles dans le dossier.

Pour les informations papier, une salle d'archives permet leur stockage. Pas de référencement des documents.

Quel est le processus de contrôle qualité des données ?

Le processus de contrôle qualité des données est en cours de rédaction par le responsable qualité de l'unité. Celui-ci intégrera notamment l'aspect validation des données.

La validation des données consiste à s'assurer que les données transmises correspondent aux attentes des partenaires des projets (protocole, lot ou expérimentation concerné), qu'elles ne comportent pas d'erreur de saisie, que les unités de mesure soient correctement transcrites, que la date et le lieu de la collecte sont facilement identifiables et si possible que l'appareil utilisé lors de la collecte apparaisse sur le fichier transmis. Elle est réalisée par le responsable du projet sur l'unité avant la transmission.

A terme une chaîne de validation sera mise en place pour affiner et fiabiliser la diffusion des données.

Quant au reste du processus, celui-ci prend en compte l'établissement d'une vérification métrologique pertinente et efficace des appareils de collecte. Les appareils de mesure (balances, pipettes) sont contrôlés par une entreprise externe une fois par an.

Stockage et sécurité des données

Quels sont les types de flux empruntés par les données et les supports utilisés pour les stocker ? (Faire éventuellement un lien vers un schéma)

Lieu et fréquence de sauvegarde :

- **Les données d'expérimentation** (numériques) sont transmises au chercheur en charge du projet soit par mail soit en main propre. Elles sont également conservées sur notre serveur. Les données non numériques (cahiers de labo, photocopie de données brutes) sont conservées sur places et numérisées. Elles sont transmises uniquement sur demande aux intervenants du projet.

- **Les données de travail** (fichier de prévisionnel de croissance, etc.) sont partiellement maintenues sur ordinateur de travail le temps de la collecte. Celles-ci sont copiées sur USB puis transférées sur le serveur et au chercheur dans un intervalle relativement court (journée).

Une petite partie des données est maintenue sur ordinateur (plus "commode") à court ou moyen terme sans sauvegarde sur serveur. Des réflexions sont en cours pour diminuer cette quantité d'informations non sécurisées (prévention, rappel des risques encourus pour les données si problème d'ordinateur, bonnes pratiques informatiques).

Quelle est la volumétrie actuelle et prévisionnelle ?

Actuellement un peu moins de 500 Go sont utilisés pour stocker les données informatiques sur notre serveur. Il y a également une salle d'archives pour les données papier.

En ce qui concerne le prévisionnel, 230 Go sont encore disponibles sur le serveur. Il est également prévu de trier et réorganiser les informations sur celui-ci pour limiter les duplicatas de fichiers et surtout de photos. Il est également en instance de définir des durées de stockage pour chaque type de données collecté et donc de prévoir la destruction des données inutilisables.

L'augmentation de la production de données audiovisuelles nécessitera d'une part une définition claire de la durée de vie de ces données et peut-être une augmentation du volume de stockage.

**L'entité hébergeant physiquement les données a-t-elle une politique de sécurité pour son système d'information ?
*politique locale, charte des infrastructures de recherche...***

[Charte des infrastructures de recherche à l'Inrae](#)

Le serveur présent sur le site n'est accessible qu'en déclarant les appareils dédiés (pas d'ordinateur/téléphone personnel) ou en utilisant le VPN mis à disposition par le service informatique. Il est accessible par l'identification des agents. Les restrictions d'accessibilité concernent particulièrement les visiteurs (chercheurs/intervenants extérieurs) et les stagiaires courte durée (inférieur à 2 mois) auxquels nous ne fournissons pas d'accès au serveur.

Les données informatiques sont sauvegardées tous les jours sur un serveur situé sur le centre INRAE de Rennes avec une durée de conservation de 30 jours.

L'accès aux données papier n'est accessible qu'avec la possession d'un badge personnel et il y a un système de surveillance et d'alarme anti-intrusion.

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 ?

- **Données expérimentales** : les acteurs ayant accès aux données et les modalités de partage sont définis par le PGD du projet notamment sur la durée d'exploitation et dans quelle mesure ces données peuvent être partagées.

- **Données d'élevage** : avec le logiciel SICPA, l'accès aux informations est accordée dès la création de l'expérimentation. Pour les données antérieures au SICPA (Ecu2000) une transmission est faite sur demande du chercheur en charge du projet par le responsable de l'expérimentation.

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

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.

- Perte d'information d'élevage, liée à une erreur de manipulation (suppression de ligne sur ECU et/ou SICPA) : pour le SICPA, contact de l'équipe informatique en charge de son développement pour récupération ; sinon, récupération de l'information de la version papier et réenregistrement ; réutilisation des photocopies de données brutes (pesée, mouvement, passage) généralement conservées dans le classeur de l'expérimentation ou du technicien en charge du lot.

- 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 ré-exporter 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.

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 ?

Les données permettant le suivi et l'évaluation des évolutions environnementales comme les températures ou les paramètres physiques et chimiques de l'eau peuvent être conservées sans limite de durée. D'une part parce que le volume de stockage qui leur est alloué est faible et d'autre part parce que l'évolution de ces données peut-être lente et peu visible d'une année sur l'autre. Cela pourrait avoir un intérêt à long terme.

Les données de généalogie doivent également être conservées sur le long terme pour retracer l'entièreté des liens de parentés du matériel biologique entretenu et sauvegardé sur l'unité.

Les données d'expérimentation produites dans le cadre de projets sont conservées en fonction de ce qui a été déterminé dans le PGD du projet en lui même.

Pour les autres données zootechniques, elles peuvent être conservées à moyen terme, mais il n'est pas pertinent de les garder plus de 20 ans après la fin des projets auxquelles elles sont associées. D'une part puisque l'évolution des appareils de mesure et des protocoles de collecte ne permet pas de maintenir une pertinence forte quant à la réutilisation de données à plus long terme et d'autre part parce qu'en l'état, l'affiliation des métadonnées aux jeux de données est incomplète. Celle-ci est insuffisante pour aller au delà d'une limite de quelques années.

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 ?

La politique de l'INRAE permet l'archivage de données à long terme sur des entrepôts sécurisés. Dans le cas de l'unité, la publication sur des entrepôts de données comme Data inrae est prévue dans un futur relativement proche pour permettre l'accessibilité et le réemploi des jeux de données.

Quelle est la durée de conservation des données ?

Pour le moment trois durées de conservation des données sont retrouvées

- Court terme : pour les données expérimentales intégrées aux projets, la conservation des données est définie dans le PGD du projet, de l'ordre de quelques années.
- Moyen terme : les données réglementaires comme les bons de livraison et les ordonnances vétérinaires ou les documents qualité sont conservés pendant 5 ans environ.
- Long terme : les autres données notamment de mesures environnementales sont conservées sans limitation de durée.

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

Les volumes de données concernés par une conservation à long terme sont assez réduits. Ils sont pour le moment couverts par la même source de financement qui permet de sauvegarder notre serveur informatique. Ces volumes sont intégrés au serveur et ne sont pas (encore) partagés sur des entrepôts de données.

Dans le cadre d'un dépôt long terme sur un entrepôt de données type Recherche data gouv, celui ci est financé par l'institution, l'INRAE, sans limitation de durée pour le moment.