



**HAL**  
open science

**Plan de gestion de données (PDG) du projet ”  
VEG&LAIT ” produire, transformer et innover à la  
ferme un système multiperformant pour relocaliser et  
diversifier l’offre alimentaire en protéines de  
légumineuses .**

Fanny Guyomarc’H

► **To cite this version:**

Fanny Guyomarc’H. Plan de gestion de données (PDG) du projet ” VEG&LAIT ” produire, transformer et innover à la ferme un système multiperformant pour relocaliser et diversifier l’offre alimentaire en protéines de légumineuses .. INRAE Bretagne Normandie. 2024. hal-04662590

**HAL Id: hal-04662590**

**<https://hal.inrae.fr/hal-04662590>**

Submitted on 26 Jul 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 Public Domain Mark 4.0 International License

# DATA MANAGEMENT PLAN (DMP) DU PROJET "VEG-ET-LAIT" "PRODUCING, PROCESSING AND INNOVATING ON THE FARM: A MULTI-PERFORMING SYSTEM FOR RELOCATING AND DIVERSIFYING THE FOOD SUPPLY OF LEGUME PROTEINS"

## PLAN DE GESTION DE DONNEES (PGD) DU PROJET « VEG&LAIT » : PRODUIRE, TRANSFORMER ET INNOVER A LA FERME : UN SYSTEME MULTIPERFORMANT POUR RELOCALISER ET DIVERSIFIER L'OFFRE ALIMENTAIRE EN PROTEINES DE LEGUMINEUSES

*Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - Modèle de PGD (français)"  
fourni par Agence nationale de la recherche (ANR).*

### Table des matières

DMP du projet "VEG-ET-LAIT "Producing, processing and innovating on the farm : a multi-performing system for relocating and diversifying the food supply of legume proteins".....	1
Renseignements sur le plan.....	5
Renseignements sur le projet .....	6
DMP du projet "VEG-ET-LAIT "Producing, processing and innovating on the farm : a multi-performing system for relocating and diversifying the food supply of legume proteins".....	8
I. ENQ_TRAQUE_INNOV : DONNEES D'ENQUETE - TRAQUE AUX INNOVATIONS.....	8
1. Description des données et collecte ou réutilisation de données existantes.....	8
2. Documentation et qualité des données .....	8
3. Stockage et sauvegarde pendant le processus de recherche.....	9
4. Exigences légales et éthiques, codes de conduite.....	9
5. Partage des données et conservation à long terme .....	10
6. Responsabilités et ressources en matière de gestion des données .....	10
II. ENQ_HISTORIQUE : DONNEES D'ENQUETE - ANALYSE SITUATIONNELLE ET HISTORIQUE .....	11
1. Description des données et collecte ou réutilisation de données existantes.....	11
2. Documentation et qualité des données .....	11
3. Stockage et sauvegarde pendant le processus de recherche.....	12
4. Exigences légales et éthiques, codes de conduite.....	12
5. Partage des données et conservation à long terme .....	13
6. Responsabilités et ressources en matière de gestion des données .....	13
III. ENQ_REFLEXIVE : DONNEES D'ENQUETE - ANALYSE REFLEXIVE IN ITINERE .....	14
1. Description des données et collecte ou réutilisation de données existantes.....	14
2. Documentation et qualité des données .....	14
3. Stockage et sauvegarde pendant le processus de recherche.....	15
4. Exigences légales et éthiques, codes de conduite.....	15
5. Partage des données et conservation à long terme .....	16
6. Responsabilités et ressources en matière de gestion des données .....	16

IV. DATA_CULTURES_LEG : DONNEES D'EXPERIMENTATIONS AGRONOMIQUES DE CULTURES DE LEGUMINEUSES .....	17
1. Description des données et collecte ou réutilisation de données existantes.....	17
2. Documentation et qualité des données .....	17
3. Stockage et sauvegarde pendant le processus de recherche.....	18
4. Exigences légales et éthiques, codes de conduite.....	18
5. Partage des données et conservation à long terme.....	19
6. Responsabilités et ressources en matière de gestion des données .....	19
V. DATA_BILANS_FERME : DONNEES DE PERFORMANCE DES FERMES (ALIMENTATION BOVINS LAIT, AZOTE, ETC) .....	20
1. Description des données et collecte ou réutilisation de données existantes.....	20
2. Documentation et qualité des données .....	20
3. Stockage et sauvegarde pendant le processus de recherche.....	21
4. Exigences légales et éthiques, codes de conduite.....	21
5. Partage des données et conservation à long terme .....	21
6. Responsabilités et ressources en matière de gestion des données .....	22
VI. SUPPORT_PARTAGE_FERM : OUTILS ET SUPPORTS POUR LE PARTAGE DE PRATIQUES VERS/ENTRE LES FERMES .....	22
1. Description des données et collecte ou réutilisation de données existantes.....	22
2. Documentation et qualité des données .....	23
3. Stockage et sauvegarde pendant le processus de recherche.....	23
4. Exigences légales et éthiques, codes de conduite.....	23
5. Partage des données et conservation à long terme .....	24
6. Responsabilités et ressources en matière de gestion des données .....	25
VII. SUPPORT_INNOV_FERM : OUTILS ET SUPPORTS POUR LA CONCEPTION INNOVANTE PAS A PAS 25	
1. Description des données et collecte ou réutilisation de données existantes.....	25
2. Documentation et qualité des données .....	25
3. Stockage et sauvegarde pendant le processus de recherche.....	26
4. Exigences légales et éthiques, codes de conduite.....	26
5. Partage des données et conservation à long terme .....	26
6. Responsabilités et ressources en matière de gestion des données .....	27
VIII. ENQ_IDEA4 : DONNEES D'ENQUETE - DIAGNOSTICS IDEA4 .....	27
1. Description des données et collecte ou réutilisation de données existantes.....	27
2. Documentation et qualité des données .....	28
3. Stockage et sauvegarde pendant le processus de recherche.....	29
4. Exigences légales et éthiques, codes de conduite.....	30
5. Partage des données et conservation à long terme .....	30
6. Responsabilités et ressources en matière de gestion des données .....	31
IX. OBJ_INGREDIENTS : GRAINES, FARINES, CONCENTRATS OU ISOLATS PRODUITS PAR LE PROJET 31	
1. Description des données et collecte ou réutilisation de données existantes.....	31
2. Documentation et qualité des données .....	32
3. Stockage et sauvegarde pendant le processus de recherche.....	32

4. Exigences légales et éthiques, codes de conduite.....	33
5. Partage des données et conservation à long terme.....	33
6. Responsabilités et ressources en matière de gestion des données .....	34
X. DATA_FAB_INGREDIENTS : DONNEES DE RECOLTE, TRI, NETTOYAGE ET SEPARATION DES INGREDIENTS VEGETAUX.....	34
1. Description des données et collecte ou réutilisation de données existantes.....	34
2. Documentation et qualité des données.....	34
3. Stockage et sauvegarde pendant le processus de recherche.....	35
4. Exigences légales et éthiques, codes de conduite.....	35
5. Partage des données et conservation à long terme.....	36
6. Responsabilités et ressources en matière de gestion des données .....	36
XI. DATA_PROP_INGREDIENTS : DONNEES ANALYTIQUES DES COMPOSITIONS ET PROPRIETES FONCTIONNELLES DES INGREDIENTS .....	37
1. Description des données et collecte ou réutilisation de données existantes.....	37
2. Documentation et qualité des données.....	37
3. Stockage et sauvegarde pendant le processus de recherche.....	38
4. Exigences légales et éthiques, codes de conduite.....	38
5. Partage des données et conservation à long terme.....	39
6. Responsabilités et ressources en matière de gestion des données .....	39
XII. OBJ_PRODUICTS : ALIMENTS MIXTES DE TYPE YAOURT, CREME OU GLACE .....	39
1. Description des données et collecte ou réutilisation de données existantes.....	39
2. Documentation et qualité des données.....	40
3. Stockage et sauvegarde pendant le processus de recherche.....	40
4. Exigences légales et éthiques, codes de conduite.....	41
5. Partage des données et conservation à long terme.....	41
6. Responsabilités et ressources en matière de gestion des données .....	42
XIII. DATA_FAB_PRODUICTS : DONNEES D'ITINERAIRES TECHNOLOGIQUES PLATEFORME LAIT .....	42
1. Description des données et collecte ou réutilisation de données existantes.....	42
2. Documentation et qualité des données.....	43
3. Stockage et sauvegarde pendant le processus de recherche.....	43
4. Exigences légales et éthiques, codes de conduite.....	44
5. Partage des données et conservation à long terme.....	44
6. Responsabilités et ressources en matière de gestion des données .....	45
XIV. DATA_PROP_PRODUICTS : DONNEES ANALYTIQUES DES COMPOSITIONS ET PROPRIETES FONCTIONNELLES DES PRODUICTS .....	45
1. Description des données et collecte ou réutilisation de données existantes.....	45
2. Documentation et qualité des données.....	46
3. Stockage et sauvegarde pendant le processus de recherche.....	47
4. Exigences légales et éthiques, codes de conduite.....	47
5. Partage des données et conservation à long terme.....	48
6. Responsabilités et ressources en matière de gestion des données .....	48
XV. DATA_CONFOCAL : IMAGES DE MICROSTRUCTURES DE PRODUICTS.....	49
1. Description des données et collecte ou réutilisation de données existantes.....	49

2. Documentation et qualité des données .....	49
3. Stockage et sauvegarde pendant le processus de recherche.....	50
4. Exigences légales et éthiques, codes de conduite.....	50
5. Partage des données et conservation à long terme .....	50
6. Responsabilités et ressources en matière de gestion des données .....	51
XVI. DATA_PROTEOLYSE : DONNEES ANALYTIQUES DE DIGESTION, PROTEOLYSE OU COMPOSITIONS EN ACIDES AMINES.....	52
1. Description des données et collecte ou réutilisation de données existantes.....	52
2. Documentation et qualité des données .....	52
3. Stockage et sauvegarde pendant le processus de recherche.....	53
4. Exigences légales et éthiques, codes de conduite.....	53
5. Partage des données et conservation à long terme .....	54
6. Responsabilités et ressources en matière de gestion des données .....	55
XVII. DATA_SAFETY : DONNEES DE CONTRÔLE MICROBIOLOGIQUE DES FABRICATIONS.....	55
1. Description des données et collecte ou réutilisation de données existantes.....	55
2. Documentation et qualité des données .....	55
3. Stockage et sauvegarde pendant le processus de recherche.....	55
4. Exigences légales et éthiques, codes de conduite.....	56
5. Partage des données et conservation à long terme .....	56
6. Responsabilités et ressources en matière de gestion des données .....	57
XVIII. DATA_FERMENTATIONS : DONNEES ANALYTIQUES DE SUIVI DES FERMENTATIONS.....	57
1. Description des données et collecte ou réutilisation de données existantes.....	57
2. Documentation et qualité des données .....	58
3. Stockage et sauvegarde pendant le processus de recherche.....	58
4. Exigences légales et éthiques, codes de conduite.....	59
5. Partage des données et conservation à long terme .....	59
6. Responsabilités et ressources en matière de gestion des données .....	60
XIX. DATA_ACV : DONNEES D'INVENTAIRES ET D'ANALYSE DU CYCLE DE VIE .....	60
1. Description des données et collecte ou réutilisation de données existantes.....	60
2. Documentation et qualité des données .....	61
3. Stockage et sauvegarde pendant le processus de recherche.....	62
4. Exigences légales et éthiques, codes de conduite.....	62
5. Partage des données et conservation à long terme .....	63
6. Responsabilités et ressources en matière de gestion des données .....	63
XX. DATA_SENSORIEL : DONNEES D'EVALUATION SENSORIELLE DES PRODUITS .....	64
1. Description des données et collecte ou réutilisation de données existantes.....	64
2. Documentation et qualité des données .....	64
3. Stockage et sauvegarde pendant le processus de recherche.....	65
4. Exigences légales et éthiques, codes de conduite.....	65
5. Partage des données et conservation à long terme .....	66
6. Responsabilités et ressources en matière de gestion des données .....	66
XXI. DATA_CONSO : DONNEES DE TESTS CONSOMMATEURS (PROJECTIVE MAPPING ET TESTS LONGITUDINAUX).....	67

1. Description des données et collecte ou réutilisation de données existantes .....	67
2. Documentation et qualité des données .....	67
3. Stockage et sauvegarde pendant le processus de recherche.....	68
4. Exigences légales et éthiques, codes de conduite.....	68
5. Partage des données et conservation à long terme .....	69
6. Responsabilités et ressources en matière de gestion des données .....	69
XXII. ENQ_MARCHE : DONNEES ET ANALYSE ETUDE DE MARCHE DES PRODUITS MIXTES .....	70
1. Description des données et collecte ou réutilisation de données existantes .....	70
2. Documentation et qualité des données .....	70
3. Stockage et sauvegarde pendant le processus de recherche.....	71
4. Exigences légales et éthiques, codes de conduite.....	71
5. Partage des données et conservation à long terme .....	71
6. Responsabilités et ressources en matière de gestion des données .....	72
XXIII. ENQ_SAFETY : DONNEES D'ENQUETE - BIBLIO RISQUES SANITAIRES.....	72
1. Description des données et collecte ou réutilisation de données existantes .....	72
2. Documentation et qualité des données .....	73
3. Stockage et sauvegarde pendant le processus de recherche.....	73
4. Exigences légales et éthiques, codes de conduite.....	73
5. Partage des données et conservation à long terme .....	74
6. Responsabilités et ressources en matière de gestion des données .....	75
XXIV. DATA_CHALLENGE_TESTS : DONNEES ANALYTIQUES DE SUIVI DES CHALLENGE TESTS .....	75
1. Description des données et collecte ou réutilisation de données existantes .....	75
2. Documentation et qualité des données .....	75
3. Stockage et sauvegarde pendant le processus de recherche.....	76
4. Exigences légales et éthiques, codes de conduite.....	76
5. Partage des données et conservation à long terme .....	77
6. Responsabilités et ressources en matière de gestion des données .....	77
XXV. DATA_HOLISTIC : DONNEES D'EVALUATION HOLISTIQUE / BALANCE BENEFICES-RISQUES .....	78
1. Description des données et collecte ou réutilisation de données existantes .....	78
2. Documentation et qualité des données .....	78
3. Stockage et sauvegarde pendant le processus de recherche.....	79
4. Exigences légales et éthiques, codes de conduite.....	79
5. Partage des données et conservation à long terme .....	80
6. Responsabilités et ressources en matière de gestion des données .....	80

## RENSEIGNEMENTS SUR LE PLAN

Titre du plan	DMP du projet "VEG-ET-LAIT "Producing, processing and innovating on the farm : a multi-performing system for relocating and diversifying the food supply of legume proteins"
Livrable	D6.1.3
Version	Version initiale
Objet/périmètre du plan	Ce plan a pour objet de désigner et de décrire les différentes données produites par les partenaires pendant le déroulé du projet ; leur nature, leur volume, et la manière dont leur transmission et leur archivage sont réalisés.

Domaines de recherche (selon classification de l'OCDE)	Agriculture, forestry, and fisheries, Animal and dairy science
Langue	fra
Date de création	2024-03-05
Date de dernière modification	2024-07-01

## RENSEIGNEMENTS SUR LE PROJET

Titre du projet	VEG-ET-LAIT "Producing, processing and innovating on the farm : a multi-performing system for relocating and diversifying the food supply of legume proteins"
Acronyme	VEG-ET-LAIT
Résumé	<p>En France, environ 6% des émissions de gaz à effet de serre proviennent de la production et de la transformation laitière. A masse égale, les boissons et « yaourts » végétaux sont environ deux fois moins émetteurs, mais ils sont aussi moins intéressants sur le plan nutritionnel et moins appréciés des consommateurs et consommatrices. Que ce soit en production ou en transformation laitière, les légumineuses offrent des leviers de diversification, d'autonomisation socio-technique et de réduction des impacts environnementaux. Au champ, les légumineuses (dont la surface a fortement diminué depuis les années 1990) ont de forts atouts agronomiques et environnementaux. Cependant, encore peu d'initiatives sont développées et coordonnées entre les deux échelons de la filière. VEG&amp;LAIT propose de jouer sur la complémentarité entre le lait et les légumineuses pour concevoir de manière intégrée une filière de produits « mixtes » associant ces deux ingrédients, de la ferme à la petite cuillère. Ce premier objectif de VEG&amp;LAIT s'inscrit pleinement dans l'axe 2 de l'appel à projets, qui concerne les enjeux de la transformation et de la consommation de légumineuses. En tant que prototype d'intégration de filière, le projet propose de s'appuyer sur des réseaux de producteurs et productrices assurant la transformation à la ferme et la distribution en circuit court. Ce dispositif permettra de mieux identifier les points de contrôle nécessaires, par exemple via des cahiers des charges, de mieux assurer la valorisation de tous les coproduits et enfin de concevoir les outils de capitalisation et de diffusion des innovations ainsi expérimentées. Ce second objectif de VEG&amp;LAIT s'inscrit quant à lui dans l'axe 3 de l'appel à projets, qui adresse les enjeux de la diversification et de l'optimisation des systèmes alimentaires. VEG&amp;LAIT est organisé en 6 work packages dont 3 sont en lien avec la construction de la filière « mixte » lait-légumineuses (WP1, WP2 et WP3) et 2 sont en lien avec la montée en compétence organisationnelle des réseaux de fermes (WP4) et l'évaluation multicritère du système ainsi constitué (WP5). Le WP6 prendra en charge le pilotage et l'animation du projet. Un état des lieux sera fait des pratiques existantes mettant en lien production et transformation des légumineuses à la ferme, dans et au-delà des réseaux impliqués dans le projet. La spécificité des fermes laitières sera prise en compte, par exemple sur le cycle de l'azote. Les fermes seront accompagnées dans la mise en place de cultures de légumineuses et dans le partage des résultats de ces expérimentations au sein du réseau. La première transformation des graines sera assurée par un partenaire expert. La mise en œuvre des farines et isolats sera réalisée sur des matrices de type « yaourt », crème dessert et crème glacée, qui valoriseront les protéines, l'amidon, les phospholipides ou des particules végétales. L'interdépendance entre la composition, les fonctionnalités des ingrédients et les propriétés finales du produit en termes de texture, de structure et d'apports nutritionnels sera finement analysée. Dans le cas des produits fermentés, la sélection des souches microbiennes capitalisera sur les connaissances de leurs propriétés d'hydrolyse pour constituer un consortium adapté. Positionnées a priori comme aliments « transitionnels », ces matrices mixtes seront utilisées pour évaluer les représentations que s'en font les consommateurs et consommatrices, selon qu'ils ou elles les compareront à des produits laitiers ou à des substituts végétaux commerciaux et ce de manière longitudinale. Dans la seconde partie du projet, une évaluation holistique de la filière ainsi déployée s'appuiera sur les informations recueillies sur les dimensions de marché, de nutrition, d'environnement, socio-techniques et sanitaires. Cette évaluation fournira des connaissances pour un accompagnement de l'innovation dans des fermes productrices et transformatrices de lait et de légumineuses.</p>
Sources de financement	<ul style="list-style-type: none"> <li>• Agence Nationale de la Recherche : ANR-23-PLÉG-0006</li> </ul>
Date de début	2024-03-01
Date de fin	2029-02-28
Partenaires	<ul style="list-style-type: none"> <li>• Science et Technologie du Lait et de l'Œuf</li> <li>• Centre des Sciences du Goût et de l'Alimentation</li> <li>• SECurité des ALiments et Microbiologie</li> <li>• Environnement, Territoires en Transition, Infrastructures, Sociétés 1456 (code codique)</li> </ul>

	<ul style="list-style-type: none"> <li>• AGRONOMIE</li> <li>• TERRES INOVIA</li> <li>• VALOREX</li> <li>• INVITATION A LA FERME</li> </ul>
--	--

## Contributeurs

Nom	Affiliation	Rôles
Arvisenet Gaëlle <a href="#">0000-0001-7994-5610</a>	UMR CSGA - 201019359D	<ul style="list-style-type: none"> <li>• Personne contact pour les données (DATA_CONSO)</li> </ul>
Carayon David <a href="#">0000-0003-0217-2175</a>	UR ETTIS - 201421784X	
Dauguet Sylvie <a href="#">0000-0002-2926-5518</a>	Terres Inovia	
Delaunay Louis <a href="#">0009-0001-4092-7041</a>	UMR SECALIM -199517939A	<ul style="list-style-type: none"> <li>• Personne contact pour les données (ENQ_SAFETY, DATA_CHALLENGE TESTS)</li> </ul>
Gagnaire Valérie <a href="#">0000-0003-3186-0620</a>	UMR STLO - 199117840K	<ul style="list-style-type: none"> <li>• Personne contact pour les données (DATA_FERMENTATIONS, DATA_PROTEOLYSE)</li> </ul>
Garric Gilles <a href="#">0000-0002-4540-2009</a>	UMR STLO - 199117840K	<ul style="list-style-type: none"> <li>• Personne contact pour les données (DATA_FAB_PRODUIITS, DATA_SAFETY)</li> </ul>
Girard Sydney <a href="#">0000-0001-5474-9699</a>	UR ETTIS	
Guérard Tiphaine	INVITATION A LA FERME	
Guyomarc-h Fanny <a href="#">0000-0002-4481-5610</a>	UMR STLO - 199117840K	<ul style="list-style-type: none"> <li>• Coordinateur de projet</li> <li>• Personne contact pour les données (DATA_FAB_INGREDIENTS, OBJ_INGREDIENTS, OBJ_PRODUIITS, DATA_PROP_INGREDNTS, DATA_ACV, DATA_PROP_PRODUIITS, DATA_CONFOCAL, ENQ_MARCHE, DATA_SENSORIEL)</li> <li>• Responsable du plan</li> </ul>
Harel-Oger Marielle 0009-0009-5992-2767	UMR STLO - 199117840K	
Jeuffroy Marie-Hélène <a href="#">0000-0003-0520-9172</a>	UMR AGRONOMIE - 196417851P	<ul style="list-style-type: none"> <li>• Personne contact pour les données (SUPPORT_INNOV_FERM)</li> </ul>
Leclère Margot 0000-0003-2212-8307	UMR AGRONOMIE - 196417851P	<ul style="list-style-type: none"> <li>• Personne contact pour les données (ENQ_HISTORIQUE, ENQ_REFLEXIVE, SUPPORT_PARTAGE_FERM)</li> </ul>
Loyce Chantal <a href="#">0000-0003-3126-8141</a>	UMR AGRONOMIE - 196417851P	
Membré Jeanne-Marie <a href="#">0000-0001-6751-4426</a>	UMR SECALIM - 199517939A	<ul style="list-style-type: none"> <li>• Personne contact pour les données (DATA_HOLISTIC)</li> </ul>
Paut Raphaël <a href="#">0000-0002-9558-6627</a>	UMR AGRONOMIE- 196417851P	<ul style="list-style-type: none"> <li>• Personne contact pour les données (DATA_BILANS_FERME, DATA_CULTURES_LEG)</li> </ul>
Tanguy Gaëlle <a href="#">0000-0001-9453-8407</a>	UMR STLO - 199117840K	
Toffolini Quentin <a href="#">0000-0002-0838-4817</a>	UMR AGRONOMIE - 196417851P	<ul style="list-style-type: none"> <li>• Personne contact pour les données (ENQ_TRAQUE_INNOV)</li> </ul>
Zahm Frédéric - <a href="#">0000-0001-6484-3128</a>	UR ETTIS- 201421784X	<ul style="list-style-type: none"> <li>• Personne contact pour les données (ENQ_IDEA4)</li> </ul>



# DMP DU PROJET "VEG-ET-LAIT "PRODUCING, PROCESSING AND INNOVATING ON THE FARM : A MULTI-PERFORMING SYSTEM FOR RELOCATING AND DIVERSIFYING THE FOOD SUPPLY OF LEGUME PROTEINS

## I. ENQ\_TRAQUE\_INNOV : DONNEES D'ENQUETE - TRAQUE AUX INNOVATIONS

### 1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

#### 1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?

La tâche 1.1 du WP1 de VEG&LAIT est abordée sous l'angle du recensement et de l'analyse de cas d'étude. De nouvelles données seront recueillies, dont :

i) Des données issues d'entretiens semi-directifs des agriculteurs et agricultrices pré-identifié-es par INVITATION A LA FERME et/ou l'association LEGGO (prestataire privilégié de VEG&LAIT).

Elles comprennent des données :

a : prioritairement qualitatives - Conduites d'entretiens qui seront menés dans le WP1, enregistrés et transcrits

b : accessoirement quantitatives – Recueil de données techniques et commerciales caractérisant l'historique de productions passées de légumineuses sur les fermes recensées

ii) Des données issues d'observations, vidéos et/ou photos.

Des observations seront conduites sur les situations de travail au sein des fermes. Ces observations pourront donner lieu au tournage de vidéos et à la prise de photos.

Des données préexistantes pourront éventuellement être réutilisées, dont :

i) Des données issues de projets antérieurs conduits par l'UMR AGRONOMIE et faisant partie de ses connaissances propres, telles que des verbatims d'agriculteur-trices ou des données techniques pertinentes par rapport au contexte des fermes étudiées, permettant d'affiner le questionnaire d'entretien. ii) Des données issues de projets connexes à VEG&LAIT et également conduits par l'UMR AGRONOMIE, tels que le projet doctoral de Maïté de Sainte Agathe.

#### 1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?

Les données sont principalement textuelles et au format .txt ou .doc ou .odt ou .pdf. Des données enregistrées seront également produites, au format .mp3, mp4 ou .wav, puis transcrites avec un logiciel créé par l'UMR AGRONOMIE utilisant la plateforme WHISPER. Des données d'image ou de vidéo pourront compléter le jeu de données, les formats sont .pdf ou .png ou .jpg ou .avi ou .mov

Le volume est estimé à priori à environ 0.5 à 1 Go par copie du jeu de données.

### 2. DOCUMENTATION ET QUALITE DES DONNEES

#### 2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?

Chaque dossier de données sera accompagné d'un fichier texte ou document (cahier de laboratoire) indiquant les métadonnées : il s'agit à minima d'identifier le contexte d'acquisition des données (auteur(s), personne interviewée, méthode(s) utilisée(s) pour acquérir ou analyser la donnée), les dates de collecte et de révision et les unités de mesure dans certains cas.

#### 2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?

L'observation d'une redondance des informations au cours du recensement des cas d'étude suggérera l'atteinte de la représentativité de l'échantillon.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des Données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans Recherche Data GOUV, espace institutionnelle Data INRAE.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### *3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Pendant l'en-cours de la tâche, il sera fait 3 copies :

- une copie locale sur le disque dur ou un HD externe de Quentin Toffolini, Raphaël Paut, Blandine Renault, Maïté de Sainte Agathe, Chantal Loyce, Marie-Hélène Jeuffroy et/ou Margot Leclère, sans cryptage particulier.
  - une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
  - une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet
- Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

#### *3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Les données sensibles identifiées sont les identités des membres des participants et participantes au projet VEG&LAIT, des volontaires recruté-es pour les entretiens dans les WP1 et WP4, et des volontaires recruté-es pour les tests consommateur-trices dans les WP2 et WP3 (voir section 4a) et les données couvertes par la propriété intellectuelle (voir section 4b).

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE.

### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

#### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses

parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

### **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

#### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Les données ou analyse d'entretiens sous leurs formes élaborées feront l'objet d'une anonymisation puis d'un partage sur la plateforme NEXTCLOUD réservée aux membres du projet, notamment à des fins d'analyse transversale entre les WP1 et WP4.

Les données ou entretiens en formats bruts seront conservé-es par les partenaires scientifiques et stocké-es sur leurs ordinateurs en respectant la nécessité de dupliquer ou tripler les lieux de stockage (cf section 3). Les publications scientifiques et techniques issues du projet seront dans la mesure du possible déposées dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les œuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

#### *5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL AgroParisTech.

#### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

Une suite bureautique suffit.

#### *5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

### **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

#### *6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Les données d'entretiens pour la traque aux innovations, l'enquête situationnelle et historique, et l'analyse réflexive des pratiques sont produites par les personnels en charge des WP1 et WP4 de VEG&LAIT, dont les leaders sont respectivement Marie-Hélène Jeuffroy/Chantal Loyce et Margot Leclère (INRAE, UMR Agronomie).

#### *6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. Les tâches des WP1 et WP4 sont programmées de 2024 à 2027 incluses (cf candidature).

La curation et la mise à disposition sur Data INRAE est estimée à 0.5 mois-personne par jeu de données.  
La préparation et la soumission d'un data paper est estimée à 1 mois-personne par jeu de données.

## II. ENQ\_HISTORIQUE : DONNEES D'ENQUETE - ANALYSE SITUATIONNELLE ET HISTORIQUE

### 1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

#### 1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?

La tâche 4.1 du WP4 de VEG&LAIT est abordée sous l'angle du recensement et de l'analyse de cas d'études situés géographiquement, temporellement et relationnellement. De nouvelles données seront recueillies, dont :

i) Des données issues d'entretiens semi-directifs des fondateur-trices de INVITATION A LA FERME, d'agriculteurs et agricultrices pré-identifié-es par INVITATION A LA FERME et/ou par l'association LEGGO (prestataire privilégié de VEG&LAIT).

Elles comprennent des données :

a : prioritairement qualitatives - Conduites d'entretiens qui seront menés dans le WP4, enregistrés et transcrits  
b : accessoirement quantitatives – Recueil de données techniques et commerciales caractérisant l'historique d'innovations passées, individuelles ou collectives

ii) Des données issues d'observations, vidéos et/ou photos

Des observations seront conduites sur les situations de travail au sein des fermes. Ces observations pourront donner lieu pourront donner lieu au tournage de vidéos et à la prise de photos.

Des données préexistantes pourront éventuellement être réutilisées, dont :

i) Des données issues de projets antérieurs conduits par l'UMR AGRONOMIE et faisant partie de ses connaissances propres, telles que des verbatims d'agriculteur-trices ou des données techniques pertinentes par rapport au contexte des fermes étudiées, permettant d'affiner le questionnaire d'entretien.

#### 1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?

Les données sont principalement textuelles et au format .txt ou .doc ou .odt ou .pdf. Des données enregistrées seront également produites, au format .mp3, mp4 ou .wav, puis transcrites avec un logiciel créé par l'UMR AGRONOMIE utilisant la plateforme WHISPER. Des données d'image ou de vidéo pourront compléter le jeu de données, les formats sont .pdf ou .png ou .jpg ou .avi ou .mov

Le volume est estimé à priori à environ 0.5 à 1 Go par copie du jeu de données.

### 2. DOCUMENTATION ET QUALITE DES DONNEES

#### 2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?

Chaque dossier de données sera accompagné d'un fichier texte ou d'un document physique (cahier de labo) indiquant les métadonnées : il s'agit à minima d'identifier le contexte d'acquisition des données (auteur(s), personne interviewée, méthode(s) utilisée(s) pour acquérir ou analyser la donnée), les dates de collecte et de révision et les unités de mesure dans certains cas. Pour les données d'enquête situationnelle et historique, les informations spatiales et temporelles seront collectées de manière détaillée et homogène.

#### 2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?

L'observation d'une redondance des informations au cours du recensement des cas d'étude suggérera l'atteinte de la représentativité de l'échantillon.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des Données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### *3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Pendant l'en-cours de la tâche, il sera fait 3 copies :

- une copie locale sur le disque dur ou un HD externe de Quentin Toffolini, Raphaël Paut, Blandine Renault, Maïté de Sainte Agathe, Marie-Hélène Jeuffroy, Chantal Loyce et/ou Margot Leclère, sans cryptage particulier..
- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

#### *3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Les données sensibles identifiées sont les identités des membres des participants et participantes au projet VEG&LAIT, des volontaires recruté-es pour les entretiens dans les WP1 et WP4, et des volontaires recruté-es pour les tests consommateur-trices dans les WP2 et WP3 (voir section 4a) et les données couvertes par la propriété intellectuelle (voir section 4b).

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE.

### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

#### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

### **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

#### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Les données ou analyse d'entretiens sous leurs formes élaborées feront l'objet d'une anonymisation puis d'un partage sur la plateforme NEXTCLOUD réservée aux membres du projet, notamment à des fins d'analyse transversale entre les WP1 et WP4.

Les données ou entretiens en formats bruts seront conservé-es par les partenaires scientifiques et stocké-es sur leurs ordinateurs en respectant la nécessité de dupliquer ou tripliquer les lieux de stockage (cf section 3).

Les publications scientifiques et techniques issues du projet seront dans la mesure du possible déposées dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les œuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

#### *5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL AgroParisTech.

#### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

Une suite bureautique suffit.

#### *5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

### **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

#### *6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Les données d'entretiens pour la traque aux innovation, l'enquête situationnelle et historique, et l'analyse réflexive des pratiques sont produites par les personnels en charge des WP1 et WP4 de VEG&LAIT, dont les leaders sont respectivement Marie-Hélène Jeuffroy/Chantal Loyce et Margot Leclère.

#### *6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. Les tâches des WP1 et WP4 sont programmées de 2024 à 2027 incluses (cf candidature).

La curation et la mise à disposition sur Data INRAE est estimée à 0.5 mois-personne par jeu de données.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne par jeu de données.



### III. ENQ\_REFLEXIVE : DONNEES D'ENQUETE - ANALYSE REFLEXIVE IN ITINERE

#### 1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

##### 1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?

Les tâches 1.3 et 4.2 du WP4 de VEG&LAIT sont abordées sous l'angle du suivi d'un ou de groupes d'individus, tout au long du projet, pour l'observation des évolutions de leurs pratiques d'innovation, à la fois sur le plan strictement culturel (introduction des légumineuses) et sur le plan organisationnel (processus collectifs de décision ou de partage d'informations). Cette analyse réflexive est étendue au processus de recherche lui-même, par l'observation des pratiques des membres du consortium et de l'évolution de ces pratiques au cours du projet. De nouvelles données seront recueillies, dont:

i) Des données issues d'entretiens semi-directifs des personnels de recherche et des personnels affiliés aux partenaires; des agriculteurs et agricultrices pré-identifiés par INVITATION A LA FERME et/ou l'association LEGGO (prestataire privilégié de VEG&LAIT)

Elles comprennent des données :

a : prioritairement qualitatives - Conduites d'entretiens qui seront menés dans le WP4, enregistrés et transcrits, à des fins de monitoring du processus d'innovation couplée production/transformation de légumineuses et du projet de recherche interdisciplinaire

b : accessoirement quantitatives – Recueil de données techniques et commerciales caractérisant l'historique de productions passées de légumineuses sur les fermes recensées

ii) Des données issues d'observations, vidéos et/ou photos.

Des observations seront conduites sur les situations de travail au sein des fermes. Ces observations pourront donner lieu pourront donner lieu au tournage de vidéos et à la prise de photos.

##### 1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?

Les données sont principalement textuelles et au format .txt ou .doc ou .odt ou .pdf. Des données enregistrées seront également produites, au format .mp3, mp4 ou .wav, puis transcrites avec un logiciel créé par l'UMR AGRONOMIE utilisant la plateforme WHISPER. Des données d'image ou de vidéo pourront compléter le jeu de données, les formats sont .pdf ou .png ou .jpg ou .avi ou .mov

Le volume est estimé à priori à environ 0.5 à 1 Go par copie du jeu de données.

#### 2. DOCUMENTATION ET QUALITE DES DONNEES

##### 2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?

Chaque dossier de données sera accompagné d'un fichier texte ou d'un document physique (cahier de labo) indiquant les métadonnées : il s'agit à minima d'identifier le contexte d'acquisition des données (auteur(s), personnes et/ou pratiques concernées par la surveillance in itinere et/ou la conduite d'atelier(s) ou de réunion(s) de partage de savoir-faire, méthode(s) utilisée(s) pour acquérir ou analyser la donnée), les dates de collecte et de révision et les unités de mesure dans certains cas.

##### 2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des Données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées

permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données

- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### *3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Pendant l'en-cours de la tâche, il sera fait 3 copies :

- une copie locale sur le disque dur ou un HD externe de Quentin Toffolini, Raphaël Paut, Blandine Renault, Maïté de Sainte Agathe, Marie-Hélène Jeuffroy, Chantal Loyce et/ou Margot Leclère, sans cryptage particulier..
- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

#### *3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Les données sensibles identifiées sont les identités des membres des participants et participantes au projet VEG&LAIT, des volontaires recruté-es pour les entretiens dans les WP1 et WP4, et des volontaires recruté-es pour les tests consommateur-trices dans les WP2 et WP3 (voir section 4a) et les données couvertes par la propriété intellectuelle (voir section 4b).

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE.

### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

#### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.



## 5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME

### 5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?

Les données ou analyse d'entretiens sous leurs formes élaborées feront l'objet d'une anonymisation puis d'un partage sur la plateforme NEXTCLOUD réservée aux membres du projet, notamment à des fins d'analyse transversale entre les WP1 et WP4.

Les données ou entretiens en formats bruts seront conservé-es par les partenaires scientifiques et stocké-es sur leurs ordinateurs en respectant la nécessité de dupliquer ou tripler les lieux de stockage (cf section 3). Les publications scientifiques et techniques issues du projet seront dans la mesure du possible déposées dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

### 5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL AgroParisTech.

### 5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?

Une suite bureautique suffit.

### 5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

## 6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES

### 6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?

Les données d'entretiens pour la traque aux innovation, l'enquête situationnelle et historique, et l'analyse réflexive des pratiques sont produites par les personnels en charge des WP1 et WP4 de VEG&LAIT, dont les leaders sont respectivement Marie-Hélène Jeuffroy/Chantal Loyce et Margot Leclère.

### 6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. Les tâches des WP1 et WP4 sont programmées de 2024 à 2027 incluses (cf candidature).

La curation et la mise à disposition sur Data INRAE est estimée à 0.5 mois-personne par jeu de données.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne par jeu de données.

## IV. DATA\_CULTURES\_LEG : DONNEES D'EXPERIMENTATIONS AGRONOMIQUES DE CULTURES DE LEGUMINEUSES

### 1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

#### 1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?

Ce produit de recherche concerne la production de graines de légumineuses en plein champ, chez les fermes identifiées par le partenaire INVITATION A LA FERME ou le prestataire LEGGO. Elle implique des données issues de mesures et observations qui seront réalisées sur le sol, les plantes et les graines récoltées, des données techniques telles que l'espèce, la variété et les données passeport de la semence; les dates et la nature des opérations techniques, les machines et les intrants employés et les débits de chantier; le rendement et la qualité de la production; des données calculées, telles que les émissions de différentes formes d'azote, de carbone ou d'eau dans l'air, l'eau ou le sol; et des données situationnelles telles que la nature du sol, sa profondeur, sa pente, sa réserve en eau, les conditions météorologiques et le précédent de culture.

Les nouvelles données produites par le projet seront complétées par des données préexistantes ou coexistantes réutilisées, telles que :

- les données de systèmes d'information géographique
- les données passeport de la semence
- les données de précédents culturaux dans la rotation, et les données de précédentes tentatives de cultures de légumineuses dans la même ferme ou d'autres fermes des réseaux de LEGGO ou INVITATION A LA FERME (cf. ENQ\_TRAQUE\_INNOVATIONS)
- les données météorologiques, pédologiques et hydrologiques produites par des tiers (Météo France, BRGM...)
- les données de compositions des intrants (fiches techniques)
- les modèles d'émission ou de bilan matière

#### 1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?

Les données sont principalement tabulées et au format .csv ou .xls ou .odt. Des données d'image ou de vidéo pourront compléter le jeu de données, les formats sont .pdf ou .png ou .jpg ou .avi ou .mov  
Le volume est estimé a priori à environ 1 à 2 Go par copie du jeu de données.

### 2. DOCUMENTATION ET QUALITE DES DONNEES

#### 2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?

Chaque dossier de données sera accompagné d'un fichier texte ou d'un document physique (cahier de laboratoire) indiquant les métadonnées : il s'agit à minima d'identifier le contexte d'acquisition des données (auteur(s), réponse recherchée, identification de la parcelle ou le l'exploitation, méthode(s) utilisée(s) pour acquérir ou analyser la donnée), les dates de collecte et de révision, et les unités de mesure dans certains cas.

#### 2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?

Lors des protocoles d'observations, d'analyses ou de mesures, des points d'étape sont prévus tels que l'inclusion de témoins ou références (par exemple temporelles), la répétition des tests, le prélèvement d'un nombre représentatif d'échantillons, et la mise en oeuvre de tests statistiques lors de l'exploitation des données.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>)..  
Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt

(dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### ***3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?***

Pendant l'en-cours de la tâche, il sera fait 3 copies :

- une copie locale sur le disque dur ou un HD externe de Quentin Toffolini, Raphaël Paut, Blandine Renault, Maïté de Sainte Agathe, Marie-Hélène Jeuffroy, Chantal Loyce et/ou Margot Leclère, sans cryptage particulier.
- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

#### ***3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?***

Les données sensibles identifiées sont les identités des membres des participants et participantes au projet VEG&LAIT, des volontaires recruté-es pour les cultures dans les WP1 et WP4, les données socio-économiques des exploitations, et les données couvertes par la propriété intellectuelle (voir section 4b).

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE. Les entrepôts de données de la Plateforme STLO sont sécurisés et stockés pendant 3 ans.

### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

#### ***4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?***

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### ***4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?***

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### ***4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?***

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

## 5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME

### 5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?

Les WP1 et WP4 s'étalent de 2024 à 2027. Les données sous leurs formes élaborées feront l'objet d'une anonymisation, d'une curation pour standardiser la désignation des fichiers et des paramètres, puis d'un partage sur la plateforme NEXTCLOUD réservée aux membres du projet, notamment à des fins d'analyse transversale entre les WP1 et WP4.

Les données ou entretiens en formats bruts seront conservé-es par les partenaires scientifiques et stocké-es sur leurs ordinateurs en respectant la nécessité de dupliquer ou tripler les lieux de stockage (cf section 3).

Les publications scientifiques et techniques issues du projet seront dans la mesure du possible déposées dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

### 5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL AgroParisTech.

### 5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?

Une suite bureautique suffit pour la lecture des données ou des sorties.

### 5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

## 6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES

### 6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?

Les données agronomiques sont collectées et/ou produites par les personnels en charge des WP1 et WP4 de VEG&LAIT, dont les leaders sont respectivement Marie-Hélène Jeuffroy/Chantal Loyce et Margot Leclère.

### 6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. Les tâches des WP1 et WP4 sont programmées de 2024 à 2027 incluses (cf candidature).

La curation et la mise à disposition sur Data INRAE est estimée à 0.5 mois-personne par jeu de données.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne par jeu de données.

## V. DATA\_BILANS\_FERME : DONNEES DE PERFORMANCE DES FERMES (ALIMENTATION BOVINS LAIT, AZOTE, ETC)

### 1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

#### 1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?

Ce produit de recherche concerne les performances des fermes ayant produit des graines de légumineuses en plein champ (cf DATA\_CULTURES\_LEG). La caractérisation de ces performances implique des données issues de mesures et observations qui seront réalisées les cultures autres que les légumineuses, des données techniques des rotations et des itinéraires complets de la ferme; des données techniques sur l'atelier bovin laitier; des données techniques sur l'atelier de transformation à la ferme; le rendement et la qualité des différentes productions; des données calculées, telles que des bilans matière; et des données socio-économiques telles que la structure de l'emploi, du revenu et les données comptables.

Les nouvelles données produites par le projet seront complétées par des données préexistantes ou coexistantes réutilisées, telles que :

- les données comptables et techniques des années précédentes
- les données météorologiques, pédologiques et hydrologiques produites par des tiers (Météo France, BRGM...)
- les données de compositions des intrants (fiches techniques)
- les modèles d'émission ou de bilan matière

#### 1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?

Les données sont principalement tabulées et au format .csv ou .xls ou .odt. Des données d'image ou de vidéo pourront compléter le jeu de données, les formats sont .pdf ou .png ou .jpg ou .avi ou .mov  
Le volume est estimé a priori à environ 1 à 2 Go par copie du jeu de données.

### 2. DOCUMENTATION ET QUALITE DES DONNEES

#### 2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?

Chaque dossier de données sera accompagné d'un fichier texte ou d'un document physique (cahier de labo) indiquant les métadonnées : il s'agit à minima d'identifier le contexte d'acquisition des données (auteur(s), réponse recherchée, identification de la parcelle ou le l'exploitation, méthode(s) utilisée(s) pour acquérir ou analyser la donnée), les dates de collecte et de révision, et les unités de mesure dans certains cas.

#### 2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?

Lors des protocoles d'observations ou de mesures, des points d'étape sont prévus tels que l'inclusion de témoins ou références (par exemple temporelles), la répétition des tests, le prélèvement d'un nombre représentatif d'échantillons, et la mise en oeuvre de tests statistiques lors de l'exploitation des données.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### *3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Pendant l'en-cours de la tâche, il sera fait 3 copies :

- une copie locale sur le disque dur ou un HD externe de Quentin Toffolini, Raphaël Paut, Blandine Renault, Maité de Sainte Agathe, Chantal Loyce, Marie-Hélène Jeuffroy et/ou Margot Leclère, sans cryptage particulier.
- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

#### *3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Les données sensibles identifiées sont les identités des membres des participants et participantes au projet VEG&LAIT, des volontaires recruté-es pour les cultures dans les WP1 et WP4, les données socio-économiques des exploitations, et les données couvertes par la propriété intellectuelle (voir section 4b).

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE. Les entrepôts de données de la Plateforme STLO sont sécurisés et stockés pendant 3 ans.

### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

#### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagés par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5Bfr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5Bfr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

### **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

#### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Les WP1 et WP4 s'étalent de 2024 à 2027. Les données sous leurs formes élaborées feront l'objet d'une anonymisation, d'une curation pour standardiser la désignation des fichiers et des paramètres, puis d'un partage sur



la plateforme NEXTCLOUD réservée aux membres du projet, notamment à des fins d'analyse transversale entre les WP1 et WP4.

Les données ou entretiens en formats bruts seront conservé-es par les partenaires scientifiques et stocké-es sur leurs ordinateurs en respectant la nécessité de dupliquer ou tripler les lieux de stockage (cf section 3). Les publications scientifiques et techniques issues du projet seront dans la mesure du possible déposées dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les œuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

#### **5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?**

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL AgroParisTech.

#### **5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?**

Une suite bureautique suffit pour la lecture des données ou des sorties.

#### **5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?**

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

### **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

#### **6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?**

Les données agronomiques sont collectées et/ou produites par les personnels en charge des WP1 et WP4 de VEG&LAIT, dont les leaders sont respectivement Marie-Hélène Jeuffroy/Chantal Loyce et Margot Leclère.

#### **6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?**

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. Les tâches des WP1 et WP4 sont programmées de 2024 à 2027 incluses (cf candidature).

La curation et la mise à disposition sur Data INRAE est estimée à 0.5 mois-personne par jeu de données.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne par jeu de données.

## **VI. SUPPORT\_PARTAGE\_FERM : OUTILS ET SUPPORTS POUR LE PARTAGE DE PRATIQUES VERS/ENTRE LES FERMES**

### **1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES**

#### **1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?**

Le produit de recherche concerne la livraison de supports permettant de disséminer les conclusions du projet en matière de culture et transformation de légumineuse, de pair à pair entre fermes et/ou par les structures ombrelles

telles que l'entreprise INVITATION A LA FERME ou l'association LEGGO, toutes deux à la tête d'un réseau de fermes. Par essence, la tâche réutilise des données préexistantes davantage qu'elle n'en produit :

i) données préexistantes issues des WP1 et WP4 au cours du projet, notamment les DATA\_CULTURES\_LEG, les DATA\_BILANS-FERME ou les résultats d'enquêtes : ENQ\_TRAQUE\_INNOV, ENQ\_HISTORIQUE, ENQ\_REFLEXIVE, ENQ\_IDEA4;

ii) données préexistantes issues des connaissances propres de UMR AGRONOMIE, d'INVITATION A LA FERME via sa certification QualiOPI ou de l'association LEGGO, pour la conception de supports.

### *1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?*

Les données sont principalement textuelles et au format .txt ou .doc ou .odt ou .pdf. Des données d'image ou de vidéo pourront compléter le jeu de données, les formats sont .pdf ou .png ou .jpg ou .avi ou .mov. A des fins de production des livrables, les données seront mises en forme de manière à pouvoir être facilement échangées avec et entre les agriculteur-trices : par exemple sous forme de diaporama .ppt ou .pdf, de poster A4 ou A0 (logigramme, fiche technique... en .ppt ou .pdf), de formulaire technique (par exemple calculateur .xls ou .odt) voire de page web grâce à l'outil open source EZ Publish utilisé par l'INRAE.

Le volume est estimé a priori à environ 1 Go par copie du jeu de données.

## **2. DOCUMENTATION ET QUALITE DES DONNEES**

### *2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?*

Chaque dossier de données sera accompagné d'un fichier texte indiquant les métadonnées : il s'agit à minima d'identifier le contexte de production des supports (auteur(s), relecteur(s), version, date). L'apposition d'une licence CC-BY sera envisagée.

### *2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?*

La qualité des données est ici la qualité des supports et outils. Elle sera validée par les pairs, en particulier l'équipe scientifique de UMR AGRONOMIE, INVITATION A LA FERME via le cahier des charges de QualiOPI, et les compétences de communication de LEGGO. Un processus de contrôle sera également mis en place avec les usagers : contrôle de la satisfaction et de la compréhension, par exemple.

## **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

### *3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Pendant l'en-cours de la tâche, il sera fait 3 copies :

- une copie locale sur le disque dur ou un HD externe de Quentin Toffolini, Raphaël Paut, Blandine Renault, Maïté de Sainte Agathe, Marie-Hélène Jeuffroy, Chantal Loyce et/ou Margot Leclère, sans cryptage particulier.
- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

### *3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Sans objet pour ce produit de recherche. La création des supports pourra cependant être protégée par une licence CC-BY.

## **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il



s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

### **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

#### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Les WP1 et WP4 s'étalent de 2024 à 2027 et la production des supports font partie des livrables de fin de tâches. Pendant l'en-cours du projet, les produits de recherche en construction feront l'objet d'un partage sur la plateforme NEXTCLOUD réservée aux membres du projet, et seront conservé-es par les partenaires scientifiques et stocké-es sur leurs ordinateurs en respectant la nécessité de dupliquer ou tripler les lieux de stockage (cf section 3).

Les supports produits, et les publications scientifiques et techniques issues du projet seront dans la mesure du possible déposées dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les œuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

#### *5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL AgroParisTech.

#### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

Les supports seront en langue française. Pour les objets électroniques, une suite bureautique suffit.

#### *5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

## 6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES

### 6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?

Les supports et outils facilitant le partage d'expériences en innovation sont produits par les personnels en charge des WP1 et WP4 de VEG&LAIT, dont les leaders sont respectivement Marie-Hélène Jeuffroy/Chantal Loyce et Margot Leclère.

### 6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. Les tâches des WP1 et WP4 sont programmées de 2024 à 2027 incluses (cf candidature).

La curation et la mise à disposition sur Data INRAE est estimée à 0.5 mois-personne par jeu de données.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne par jeu de données.

## VII. SUPPORT\_INNOV\_FERM : OUTILS ET SUPPORTS POUR LA CONCEPTION INNOVANTE PAS A PAS

### 1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

#### 1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?

Le produit de recherche concerne la livraison de supports permettant de disséminer les conclusions du projet en matière de culture et transformation de légumineuse, de pair à pair entre fermes et/ou par les structures ombrelles telles que l'entreprise INVITATION A LA FERME ou l'association LEGGO, toutes deux à la tête d'un réseau de fermes. Par essence, la tâche réutilise des données préexistantes davantage qu'elle n'en produit :

- i) données préexistantes issues des WP1 et WP4 au cours du projet, notamment les ENQ\_TRAQUE\_INNOV, ENQ\_HISTORIQUE et ENQ\_REFLEXIVE ;
- ii) données préexistantes issues des connaissances propres de UMR AGRONOMIE, d'INVITATION A LA FERME via sa certification QualiOPI ou de l'association LEGGO, pour la conception de supports.

#### 1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?

Les données sont principalement textuelles et au format .txt ou .doc ou .odt ou .pdf. Des données d'image ou de vidéo pourront compléter le jeu de données, les formats sont .pdf ou .png ou .jpg ou .avi ou .mov. A des fins de production des livrables, les données seront mises en forme de manière à pouvoir être facilement échangées avec et entre les agriculteur-trices : par exemple sous forme de diaporama .ppt ou .pdf, de poster A4 ou A0 (logigramme, fiche technique... en .ppt ou .pdf), de formulaire technique (par exemple calculateur .xls ou .odt) voire de page web grâce à l'outil open source EZ Publish utilisé par l'INRAE.

### 2. DOCUMENTATION ET QUALITE DES DONNEES

#### 2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?

Chaque dossier de données sera accompagné d'un fichier texte indiquant les métadonnées : il s'agit à minima d'identifier le contexte de production des supports (auteur(s), relecteur(s), version, date). L'apposition d'une licence CC-BY sera envisagée.

#### 2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?

La qualité des données est ici la qualité des supports et outils. Elle sera validée par les pairs, en particulier l'équipe scientifique de UMR AGRONOMIE, INVITATION A LA FERME via le cahier des charges de QualiOPI, et les compétences de communication de LEGGO. Un processus de contrôle sera également mis en place avec les usagers : contrôle de la satisfaction et de la compréhension, par exemple.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### *3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Pendant l'en-cours de la tâche, il sera fait 3 copies :

- une copie locale sur le disque dur ou un HD externe de Quentin Toffolini, Raphaël Paut, Blandine Renault, Maité de Sainte Agathe, Marie-Hélène Jeuffroy, Chantal Loyce et/ou Margot Leclère, sans cryptage particulier.
- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

#### *3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Sans objet pour ce produit de recherche.

### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

#### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE <https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFR%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFR%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

### **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

#### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Les WP1 et WP4 s'étalent de 2024 à 2027 et la production des supports font partie des livrables de fin de tâches. Pendant l'en-cours du projet, les produits de recherche en construction feront l'objet d'un partage sur la plateforme NEXTCLOUD réservée aux membres du projet, et seront conservé-es par les partenaires scientifiques et stocké-es sur leurs ordinateurs en respectant la nécessité de dupliquer ou tripliquer les lieux de stockage (cf section 3).

Les supports produits, et les publications scientifiques et techniques issues du projet seront dans la mesure du possible déposées dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une

science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

#### *5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL AgroParisTech.

#### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

Les supports seront en langue française. Pour les objets électroniques, une suite bureautique suffit.

#### *5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

### **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

#### *6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Les supports et outils facilitant le partage d'expériences en innovation sont produits par les personnels en charge des WP1 et WP4 de VEG&LAIT, dont les leaders sont respectivement Marie-Hélène Jeuffroy/Chantal Loyce et Margot Leclère.

#### *6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. Les tâches des WP1 et WP4 sont programmées de 2024 à 2027 incluses (cf candidature).

La curation et la mise à disposition sur Data INRAE est estimée à 0.5 mois-personne par jeu de données.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne par jeu de données.

## **VIII. ENQ\_IDEA4 : DONNEES D'ENQUETE - DIAGNOSTICS IDEA4**

### **1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES**

#### *1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?*

Ce produit de recherche concerne l'évaluation des performances de durabilité agroécologique, socio-territoriale et économique des fermes ayant produit ou non des graines de légumineuses en plein champ (cf DATA\_CULTURES\_LEG), avant et après l'introduction de cette nouvelle culture.

La caractérisation de ces performances implique la collecte de données socio-techniques au cours d'entretiens directs guidés par le questionnaire IDEA4 ([https://methode-idea.org/fileadmin/user\\_upload/Documents/2.outils\\_methode/Document\\_de\\_collecte\\_des\\_donnees\\_2023\\_07\\_04.pdf](https://methode-idea.org/fileadmin/user_upload/Documents/2.outils_methode/Document_de_collecte_des_donnees_2023_07_04.pdf)). Les données collectées caractérisent le profil, la situation et l'historique de l'exploitation, et détaillent ses ressources, pratiques et résultats. Au cours de ces entretiens et du travail d'analyse exercé par l'enquêteur-trice, une valeur est déterminée quantitativement pour chacun des 53 indicateurs de la méthode IDEA4 (<https://methode-idea.org/outils/calculateur-idea4-et-outils-denquete>). L'analyse de ces résultats produit une nouvelle typologie de

données, correspondant à une évaluation de la durabilité de l'exploitation sous deux angles de lecture : analyse par les 5 propriétés (ancrage territorial, robustesse, capacité productive et reproductive, responsabilité globale et autonomie) ou par les 3 dimensions de durabilité agroécologique, socio-territoriale et économique.

Les données seront collectées chez les agriculteurs enquêtés (une dizaine) puis enregistrées sur un calculateur Excel pour le calcul des indicateurs et enfin les données de résultats seront téléversées sur la plateforme WEB IDEA4 (<https://web-idea.inrae.fr/>) par les enquêteurs

Ainsi, les données nouvelles produites sont les résultats d'analyse, téléversés sur WEB IDEA4.

Les données préexistantes réutilisées sont les outils utilisés pour l'enquête et données collectées pour l'analyse, à savoir :

- les outils et le cadre conceptuel de la méthode IDEA4 : guide, formulaire et calculateur (<https://methode-idea.org/>)
- les documents généraux demandés à l'exploitant : plans, liste des biens meubles et immeubles, relevés de consommations, déclarations PAC, etc
- les documents techniques relatifs aux différents ateliers de production et de transformation
- les documents socio-économiques relatifs à l'emploi et à la comptabilité de l'exploitation
- les résultats d'analyses IDEA4 de fermes enquêtées antérieurement au projet et dont les données peuvent être utilisées pour compléter l'échantillonnage (par exemple pour constituer un profil de référence de fermes sans production de légumineuses et/ou sans atelier de transformation à la ferme).

### 1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?

Les données IDEA4 reportées dans le fichier .json du calculateur sont toutes listées de manière exhaustive dans l'onglet spécifique dédié (« renvoi BDD ») du calculateur Excel IDEA4. Elles sont rappelées ci-dessous selon leur type d'information :

- Type 1 : données correspondant aux valeurs numériques des 119 items calculés par la biais de la méthode IDEA4 pour l'exploitation agricole enquêtée (valeurs chiffrées allant de -6 à 20, exprimées en « unités de durabilité ») (exemple pour l'item 1 de l'indicateur A1 : Diversité et répartition équilibrée des cultures, la valeur importée est une valeur entre 0 et 5).
- Type 2 : méta données correspondant aux données générales et structurelles de l'exploitation collectées et importées dans la plateforme WEB-IDEA classées en trois catégories (catégorie 1 : générale, catégorie 2 : données structurelles, et catégorie 3 : donnée individuelle).

Les données sont principalement tabulées et exportables au format .csv ou .xls ou .odt. Des données aux formats image ou vidéo pourront compléter le jeu de données, les formats sont .pdf ou .png ou .jpg ou .avi ou .mov

Le volume est estimé a priori à environ 1 à 2 Go par copie du jeu de données.

## 2. DOCUMENTATION ET QUALITE DES DONNEES

### 2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?

Le formulaire d'enquête comprend une première section de métadonnées explicitement demandées à l'exploitant enquêté ([https://methode-idea.org/fileadmin/user\\_upload/Documents/2.outils\\_methode/Document\\_de\\_collecte\\_des\\_donnees\\_2023\\_07\\_04.pdf](https://methode-idea.org/fileadmin/user_upload/Documents/2.outils_methode/Document_de_collecte_des_donnees_2023_07_04.pdf)).

Chaque dossier de données collectées pour une exploitation (formulaire et pièces documentaires) sera accompagné d'un fichier texte indiquant les métadonnées : il s'agit des métadonnées suivantes listées au tableau ci-dessous

N°	Nom	Type / étendue
1	Version du calculateur utilisée	Alpha 10
2	Identifiant exploitation	Num 6
3	SAU	0 à 10 000 (2 dc)
4	UTH	0 à 10 000 (2 dc)
5	UTH F	0 à 100 (2 dc)
6	Tranche d'âge du chef d'exploitation	« -25 », « 25-35 », « 35-45 », « 45-55 », « 55-65 » et « 65+ »
7	Typologie d'exploitation (OTEX)	Codes OTEX simplifiés

8	Surface en herbe en % de la SAU	0 à 100 (2 dc)
9	Capital d'exploitation	0 à 100 000 000
10	EBE	- 1 000 000 à 10 000 000
11	Résultat courant	- 1 000 000 à 10 000 000
12	Zone géographique (département)	liste des départements
13	Atelier hors sol : oui / non	0 ou 1
14	Année d'enquête	Num 4
15	Type d'élevage	0 - pas d'élevage / 1 - monogastrique / 2 - herbivore
16	Part des PP dans la SAU	0 à 100 (2 dc)
17	Usage des produits phytos et/ou veto : oui /non	0 ou 1
18	Exploitation de lycée agricole ou école	0 ou 1

L'identifiant calculateur est un code à 6 chiffres associé à un calculateur. Il est choisi par l'opérateur qui réalise le diagnostic et dépose le fichier .json correspondant sur la plateforme WEB-IDEA. Il permet de garantir la pseudonymisation des données rattachées à ce même calculateur. Une fois déposée dans la plateforme WEB-IDEA, les données issues du calculateur IDEA4 seront identifiées uniquement au moyen de ce code de manière strictement non nominative.

## 2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?

Les travaux du Comité scientifique de la méthode IDEA4 ont en charge la révision périodique de la méthode et la qualification des enquêteur-trices garantissent la qualité des données est assurée par des formations délivrée soit par l'équipe projet soit par le centre écodéveloppement de Villarceaux.

La tâche 5.3 est abordée avec une approche par cas d'étude. Pour dégager certaines réponses, telles que l'effet de la transformation à la ferme et/ou l'effet d'introduire des légumineuses "food" sur la durabilité globale, un nombre de cas d'étude supérieur à 4 est recherché pour chaque modalité.

## 3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE

### 3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?

Pendant l'en-cours de la tâche, il sera fait 3 copies :

- une copie locale sur le disque dur ou un HD externe de Sydney Girard, David Carayon ou Frédéric Zahm, sans cryptage particulier.

- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE

- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Les données anonymisées de résultats seront enfin téléversées sur la plateforme WEB IDEA4 <https://web-idea.inrae.fr/> par les enquêteurs

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

### 3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?

Les données sensibles identifiées sont les identités des membres des exploitations agricoles ainsi que les données sociales et comptables des exploitations engagées dans la tâche 5.3 du projet VEG&LAIT, directement (identification et recrutement par INVITATION A LA FERME ou LEGGO) ou indirectement (résultats de diagnostics antérieurs disponibles dans la base de données IDEA et utilisés pour compléter l'échantillonnage). Les données contenues dans la plateforme WEB-IDEA sont anonymisées directement par l'utilisateur et les informations (18 métadonnées ou données de résultats) contenues dans la plateforme ne permettent pas d'identifier l'exploitation agricole.

Voir aussi section 4.

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE.



## 4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE

### 4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?

Les données à caractère personnel identifiées sont les identités des exploitations agricoles volontaires recrutées pour les entretiens. Le formulaire dédié spécifiquement d'information et de consentement sera remis, conformément à l'obligation qui est faite lors de l'usage de la méthode IDEA4

Un formulaire spécifiquement dédié est téléchargeable sur le site de la méthode

[https://methode-](https://methode-idea.org/fileadmin/user_upload/Documents/10.CasdarAction/Livrables/Formulaire_consentement_destination_agriculteurs_INTERACTIF_2023_07_26.pdf)

[idea.org/fileadmin/user\\_upload/Documents/10.CasdarAction/Livrables/Formulaire\\_consentement\\_destination\\_agriculteurs\\_INTERACTIF\\_2023\\_07\\_26.pdf](https://methode-idea.org/fileadmin/user_upload/Documents/10.CasdarAction/Livrables/Formulaire_consentement_destination_agriculteurs_INTERACTIF_2023_07_26.pdf).

Ce formulaire sera actualisé en fonction des besoins et des évolutions de la réglementation.

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE et tous les utilisateurs signent les CGU préalables au téléversement des données sur la plateforme

### 4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

### 4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

## 5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME

### 5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?

La tâche 5.3 est réalisée sur la période 2024 à 2027. Les données sensibles ne sont pas partagées. Après anonymisation et analyse, les résultats sont téléversés sur la plateforme WEB IDEA4 (<https://web-idea.inrae.fr/>) par les enquêteurs.

Les publications scientifiques et techniques, ainsi que les rapports de stage issus du projet seront dans la mesure du possible déposés dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

### 5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?

La conservation et le partage de données sont inhérents à la méthode IDEA4 et se matérialisent par l'alimentation de la plateforme WEB IDEA4 (<https://web-idea.inrae.fr/>).

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL AgroParisTech.

### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

La méthode restitue des données tabulées et un rapport. Une suite bureautique suffit.

### *5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

## **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

### *6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

En tant que partenaires, l'unité ETTIS et en particulier Sydney Girard, Frédéric Zham et David Carayon sont garants de la mise en oeuvre de la méthode pour les besoins du projet et de la gestion des données issues de cette activité, y compris celles produites par le biais d'une prestation de l'UMR SAS (Institut Agro Rennes Angers) ou de Cœur Fenouil (entreprise dirigée par Sydney Girard) si cette dernière est sollicitée.

### *6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. La tâche est programmée de 2024 à 2027 incluses (cf candidature).

La collecte des données, l'analyse, la curation et le téléversement sur la plateforme IDEA4 sont estimées à 1 mois-personne par jeu de données (4 ou 5 fermes).

La curation et la mise à disposition sur Data INRAE est estimée à 0.5 mois-personne par jeu de données.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne par jeu de données.

## **IX. OBJ\_INGREDIENTS : GRAINES, FARINES, CONCENTRATS OU ISOLATS PRODUITS PAR LE PROJET**

### **1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES**

#### *1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?*

Des ingrédients de légumineuses seront nécessaires pour la conduite de VEG&LAIT. Il peut s'agir de farines, de concentrats ou d'isolats protéiques, mais aussi de fractions enrichies en amidon ou en fibre, pour anticiper un objectif de valorisation de toute la graine.

Au cours de la ou des deux premières années, des ingrédients commerciaux seront mobilisés. Le sourcing se fera auprès du prestataire LEGGO, des partenaires VALOREX et INRAE, notamment pour ce dernier via les projets en cours LETSPROSEED et DATAVEG. Les espèces prioritaires sont la féverole, le pois, la lentille, le lupin ou le pois chiche.

Dans la suite du projet, des ingrédients seront produits par les partenaires du projet, à partir de cultures conduites par INVITATION A LA FERME et transformées par VALOREX. En première intention, 6 à 8 ingrédients protéiques seront produits par mouture et fractionnement en voie sèche. Les coproduits riches en amidon ou en fibres seront également récupérés.

#### *1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?*

Achat ou production d'environ 10 kg de chaque ingrédient cible.



## 2. DOCUMENTATION ET QUALITE DES DONNEES

### 2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?

Le nom de l'espèce, de la variété ou cultivar, l'année de production, les coordonnées et la localisation du producteur, le bilan matière (quantité de graines et différentes fractions) et analyse de composition (au minimum extrait sec et teneur en protéine et/ou amidon et/ou fibres) seront répertoriés dans un fichier Excel et affichées sur les échantillons, y compris après division ou reconditionnement.

### 2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?

La qualité technique des données sera revue par les pairs TERRES INOVIA, VALOREX ou UMR AGRONOMIE ou par le prestataire LEGGO. Il s'agira notamment d'identifier les métadonnées nécessaires à la qualité de la données, telles que mentionnées en section 2a.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

## 3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE

### 3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?

Il sera fait au moins 3 copies :

- une copie locale sur le disque dur ou un HD externe de la coordinatrice ou des membres du consortium, sans cryptage particulier.
- une copie sur l'espace alloué à la coordinatrice sur le DataCenter INRAE
- une copie sur l'espace partagé dédié au projet

L'étiquetage physique des produits sera également réalisé, avec une duplication en cas de division ou reconditionnement.

### 3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?

Les échantillons d'ingrédients protéiques (et coproduits associés) seront conditionnés en sac ou boîte opaque comportant 2 étiquettes d'identification et de rattachement au projet. Ils seront partagés et stockés au sein de l'UMR STLO, d'INVITATION A LA FERME et/ou de VALOREX en chambre froide ou en cave (T<20°C) pendant la durée de l'étude.

Un inventaire des stocks sera réalisé tous les 6 mois.

Dans le cas d'ingrédients protéiques faisant partie des connaissances propres de l'un ou l'autre des partenaires du consortium, le partage doit se faire aux termes de l'accord de consortium (formulaire de demande et enregistrement daté de la concession).

#### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

##### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

##### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagés par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

##### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

#### **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

##### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Les échantillons produits dans le cadre du projet sont disponibles dès leur fabrication et rendu disponibles aux partenaires selon les quantités nécessaires aux expérimentations prévues dans le projet.

Les échantillons faisant partie des connaissances propres de l'un ou l'autre des partenaires peuvent être partagés dans le respect des termes de l'accord de consortium.

Il n'y a pas de raison de définir un embargo.

Il pourra être mis en place un contrôle périodique de la qualité microbiologique des ingrédients, afin de restreindre leur distribution si une contamination est avérée.

##### *5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Tous les ingrédients seront conservés sans sélection, sauf contamination avérée. Dans ce cas, ils seront détruits par autoclavage et élimination par la filière de traitement des ordures ménagères (incinération pour Rennes et agglomération).

##### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

sans objet

##### *5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

sans objet

## 6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES

*6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Chaque partenaire sera responsable du référencement et de la conservation des ingrédients pour la part qui lui échoit.

*6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

sans objet

## X. DATA\_FAB\_INGREDIENTS : DONNEES DE RECOLTE, TRI, NETTOYAGE ET SEPARATION DES INGREDIENTS VEGETAUX

### 1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

*1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?*

Le produit de recherche concerne spécifiquement les données de fabrications d'ingrédients (OBJ\_INGREDIENTS) réalisés par les partenaires du projet, à partir de cultures conduites par INVITATION A LA FERME et transformées par VALOREX. En première intention, 6 à 8 ingrédients protéiques seront produits par mouture et fractionnement en voie sèche. Les coproduits riches en amidon ou en fibres seront également récupérés.

En vue de maîtriser le processus conduisant à la fabrication des OBJ\_INGREDIENTS et de conduire les inventaires du cycle de vie (DATA\_ACV), les données suivantes sont nécessaires :

- Le nom de l'espèce, de la variété ou cultivar, l'année de production, les coordonnées et la localisation du producteur et tout autre donnée agronomique utile à la traçabilité ou à l'interprétation des qualités fonctionnelles de l'ingrédient
- les différentes opérations unitaires mobilisées et leur enchaînement (description de l'itinéraire technologique)
- les différentes matières relatives à la graine, entrantes (récolte) et sortantes (refus de tri, pellicules, farine, son, concentré, amidon etc), et pour chaque flux : les masses, les compositions, la destination, et le cas échéant le prix de chaque matière sortante si elle n'est pas un déchet
- Les éventuels transports : matériau et masse transportée, distance, modalité, énergie, ambiance contrôlée ou non, volume occupé/disponible
- Les éventuels stockages : matériau et masse stockée, durée, énergie, ambiance contrôlée ou non, volume occupé/disponible

Des données préexistantes pourront être réutilisées. Il s'agira par exemple :

- d'itinéraires techniques et de bilans matières issus de la littérature et/ou de dires d'experts
- des informations qui auront pu être recueillies sur les ingrédients commerciaux utilisés en première année de VEG&LAIT, notamment les compositions, l'itinéraire technique (autant que possible) et les informations de production (année, espèce, etc).

*1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?*

Les données sont principalement tabulées et au format .csv ou .xls ou .odt. Des données aux formats image ou vidéo pourront compléter le jeu de données, les formats sont .pdf ou .png ou .jpg ou .avi ou .mov

Le volume est estimé à priori à environ 1 à 2 Go par copie du jeu de données.

### 2. DOCUMENTATION ET QUALITE DES DONNEES

*2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?*

Chaque dossier de données sera accompagné d'un fichier texte ou d'un document physique (cahier de labo) indiquant les métadonnées : il s'agit à minima d'identifier le contexte d'acquisition des données (auteur(s), opérateur(s), méthode(s) utilisée(s) pour acquérir ou analyser la donnée), d'identifier la graine entrante (espèce, variété/cultivar, origine/identification de la production, lot...), d'identifier l'itinéraire technologique; d'identifier les dates de collecte et de révision des données de fabrication de l'ingrédient, et les unités de mesure.

## **2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?**

Lors des protocoles d'observations, d'analyses ou de mesures, des points d'étape sont prévus tels que le calibrage des instruments, l'inclusion de témoins ou références (par exemple temporelles), la répétition des tests, le prélèvement d'un nombre représentatif d'échantillons, et la mise en oeuvre de tests statistiques lors de l'exploitation des données.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

## **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

### **3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?**

Pendant l'en-cours de la tâche, il sera fait 3 copies :

- une copie locale sur le disque dur ou un HD externe de Fanny Guyomarc'h, Gilles Garric, Mathieu Guillevic ou Gabrielle Jagorel (CDD sous la responsabilité de F. Guyomarc'h) , sans cryptage particulier.
- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

### **3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?**

Les données sensibles identifiées sont les identités des membres des participants et participantes au projet VEG&LAIT (voir section 4a) et les données couvertes par la propriété intellectuelle (voir section 4b), notamment ici les données industrielles produites par VALOREX.

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE. Les entrepôts de données de la Plateforme STLO sont sécurisés et stockés pendant 3 ans.

## **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

### **4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?**

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

*4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagés par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

*4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

## **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

*5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Les données de DATA\_FAB\_INGREDIENTS sont par essence des données confidentielles. La maîtrise de la préparation d'ingrédients végétaux est une connaissance propre à VALOREX. Si les OBJ\_INGREDIENTS qui seront produits sont un résultat commun du projet, l'itinéraire technologique lui-même reste un résultat propre de VALOREX, qu'il pourra partager sur demande aux partenaires (par exemple à des fins d'ACV ou de mise en oeuvrens des OBJ\_PRODUIITS) mais dont il pourra empêcher la diffusion publique.

*5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Les données à conserver sont le schéma du système, les données relatives à la graine, aux opérations unitaires et aux bilans matières. Une copie du jeu de données sera conservée sur un disque personnel, une autre sur l'espace alloué sur le datacenter, et une enfin sur Data INRAE; sous réserve des dispositions de l'accord de consortium.

*5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

Une suite bureautique suffit.

*5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

Cette attribution n'est pas envisagée pour ce produit de recherche, en raison de la confidentialité des opérations.

## **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

*6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Fanny Guyomarc'h en tant que responsable du WP2.

*6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

La curation et la mise à disposition sur Data INRAE est estimée à 0.5 mois-personne.  
La préparation et la soumission d'un data paper est estimée à 1 mois-personne.

# XI. DATA\_PROP\_INGREDIENTS : DONNEES ANALYTIQUES DES COMPOSITIONS ET PROPRIETES FONCTIONNELLES DES INGREDIENTS

## 1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

### 1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?

Le produit de recherche concerne la caractérisation de la composition et des propriétés fonctionnelles des ingrédients, commerciaux ou autoproduits (OBJ\_INGREDIENTS). Les données sont typiquement obtenues par la mesure et l'analyse quantitative des réponses suivantes :

- composition en nutriments (extrait sec, protéines, lipides, amidon, cendres, fibres)
- profil granulométrique, masse volumique, couleur de la poudre
- dispersibilité et mouillabilité de la poudre dans l'eau ou une solution proxy de la phase aqueuse du lait
- solubilité en fonction du pH, de la température, et/ou de la présence de caséines, et/ou du traitement thermique de la suspension
- propriétés de gélification thermique ou acide en fonction des mêmes facteurs
- stabilité des gels dans le temps (observation de synérèse ou de décantation)

Des données préexistantes pourront être réutilisées, telles que les données fournisseur ou les données issues du projet LEGALIM partagé par VALOREX et STLO depuis 2020.

### 1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?

Les données sont principalement tabulées et au format .csv ou .xls ou .odt. Des données aux formats image ou vidéo pourront compléter le jeu de données, les formats sont .pdf ou .png ou .jpg ou .avi ou .mov  
Le volume est estimé a priori à environ 2 Go par copie du jeu de données.

## 2. DOCUMENTATION ET QUALITE DES DONNEES

### 2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?

Sauf dans le cas des analyses commandées en prestation, la réalisation de ces analyses est une expertise de l'unité STLO qui en détient le savoir-faire. La méthodologie a fait l'objet de publications antérieures, en particulier dans la thèse DeciBELL (Canon et al., 2020; 2021).

STLO s'est doté d'un système de management de la qualité et la plupart de ses modes opératoires ont fait l'objet de la rédaction d'un protocole validé et stocké dans une base de données interne (LabCollector).

Les métadonnées qui accompagneront les données seront typiquement le ou les auteur-trice-s, le rattachement au présent projet par le numéro du projet (23-010), le titre ou numéro de la collecte, la date de collecte et de ses éventuelles versions (pooling, post-traitement). Dans l'encours du projet, ces métadonnées seront consignées dans le cahier de laboratoire et dans les fiches de liaisons entre services (par exemple formulaires de demande d'analyse). En fin de projet, ces métadonnées seront plus précisément définies dans la collection "VEG&LAIT" rattachée au dataverse de STLO sur Data INRAE.

### 2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?

L'analyse des compositions, l'analyse des propriétés physiques des poudres et l'analyse des propriétés fonctionnelles des produits font l'objet de fiches Assurance Qualité en cours de rédaction au STLO. L'analyse de l'extrait sec ou de certains macronutriments, tels que les protéines, sont adossées à des normes ISO.

Lors des protocoles d'observations, d'analyses ou de mesures, des points d'étape sont prévus tels que le calibrage des instruments, l'inclusion de témoins ou références, la répétition des tests, le prélèvement d'un nombre représentatif d'échantillons, et la mise en oeuvre de tests statistiques lors de l'exploitation des données.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies



- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### *3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Pendant l'en-cours de la tâche, il sera fait 3 copies :

- une copie locale sur le disque dur ou un HD externe de Fanny Guyomarc'h, Gilles Garric ou Gabrielle Jagorel (CDD sous la responsabilité de F. Guyomarc'h) , sans cryptage particulier.
  - une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
  - une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet
- Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

#### *3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Les données sensibles identifiées sont les identités des membres des participants et participantes au projet VEG&LAIT, des volontaires recruté-es pour les entretiens dans les WP1 et WP4, et des volontaires recruté-es pour les tests consommateur-trices dans les WP2 et WP3 (voir section 4a) et les données couvertes par la propriété intellectuelle (voir section 4b).

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE. Les entrepôts de données de la Plateforme STLO sont sécurisés et stockés pendant 3 ans.

### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

#### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTologiques RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

## 5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME

### 5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?

Le WP2 s'échelonne de 2024 à 2027.

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

Les publications scientifiques et techniques, ainsi que les rapports de stage issus du projet seront dans la mesure du possible déposés dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

### 5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL Institut Agro.

### 5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?

La plupart des travaux du WP2 se tiendront à UMR STLO, qui utilise LabCollector pour le partage et l'archivage des méthodes expérimentales et de leurs révisions. Dans l'en-cours du projet, les méthodes et les réalisations seront également enregistrées et datées dans un cahier de laboratoire et un classeur physiques (papier) à défaut de numérisation. Dans l'exécution des manip, les logiciels propriétaires des machines seront utilisés in situ, mais comportent tous la possibilité d'exporter les données tabulées (exemple .csv) utilisables avec Excel.

### 5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

## 6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES

### 6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?

Fanny Guyomarc'h en tant que responsable du WP2.

### 6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?

La curation et la mise à disposition sur Data INRAE est estimée à 0.5 mois-personne.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne.

## XII. OBJ\_PRODUICTS : ALIMENTS MIXTES DE TYPE YAOURT, CREME OU GLACE

### 1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES



### *1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?*

Des prototypes de 3 types de produits mixtes (analogues de yaourt, de crème dessert ou de crème glacée) seront produits par les partenaires INRAE (Plateforme Lait de STLO), VALOREX et INVITATION A LA FERME au sein du WP2 de VEG&LAIT. Il y aura au minimum 2 prototypes par type de produits, variant principalement par la recette (nature et proportion de l'ingrédient végétal) et éventuellement par les paramètres des itinéraires technologiques. La production de ces prototypes devra mettre en œuvre des OBJ\_INGREDIENTS correspondant à des données sélectionnées en amont de cette tâche dans le projet et également référencées dans ce PGD. Le produit fermenté, analogue de yaourt, mettra en œuvre des ferments lactiques d'origine commerciale.

### *1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?*

Les fabrications seront typiquement réalisées sur des masses de recette de 5 à 50 kg, et des fabrications de 1 à 10 kg en Thermomix ou turbine à glace (volumes 3-5 L) ou sur la Plateforme Lait (10 L minimum de produit).

## **2. DOCUMENTATION ET QUALITE DES DONNEES**

### *2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?*

Pour les essais qui seront réalisés sur la Plateforme Lait du STLO, le système de management de la qualité propre à ce partenaire sera utilisé. Il met en œuvre le numéro d'action 23-010 et la liasse documentaire propre à la commande de fabrication (fiche projet) et d'analyses (fiches dédiées ; par exemple pour les analyses de composition).

### *2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?*

La Plateforme Lait applique un plan de maintenance qui assure le bon fonctionnement et la propreté des locaux et des équipements. Les principaux outils de mesure (thermomètres, pesons, débitmètres, pHmètres...) sont vérifiés au moins annuellement, et au début de chaque fabrication pour le pH et les masses. Chaque fabrication est répétée au moins 3 fois de bout en bout.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>). Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

## **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

### *3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Pendant l'en-cours de la tâche, il sera fait 3 copies :

- une copie locale sur le disque dur ou un HD externe de Fanny Guyomarc'h, Gilles Garric ou Gabrielle Jagorel (CDD sous la responsabilité de F. Guyomarc'h) , sans cryptage particulier.
- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

### *3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Les échantillons seront sous la responsabilité des partenaires industriels et pourront être mis à disposition des autres partenaires du projet pour analyses comme prévu dans le projet. En cas d'analyse sensorielle ou autres tests biologiques réalisés en interne ou par d'autres partenaires du projet, chaque partenaire sera garant de l'innocuité des produits par la mise en œuvre d'un plan de maîtrise des risques biologiques et devra assurer une logistique d'acheminement jusqu'au partenaire prévenant toute altération des échantillons.

## **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITÉES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

## **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Les produits issus du projet sont disponibles dès leur fabrication et rendu disponibles aux partenaires selon les quantités nécessaires aux expérimentations prévues dans le projet.

Les produits faisant partie des connaissances propres de l'un ou l'autre des partenaires peuvent être partagés dans le respect des termes de l'accord de consortium.

Il n'y a pas de raison de définir un embargo.

Il pourra être mis en place un contrôle périodique de la qualité microbiologique des produits, afin de restreindre leur distribution si une contamination est avérée.

### *5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Congélation si besoin, avec étiquetage complet (date, numéro de projet, composition, personne contact) y compris en cas de division ou de reconditionnement.

*5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

sans objet

*5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

sans objet

**6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

*6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Chaque partenaire sera responsable du référencement et de la conservation des produits pour la part qui lui échoit.

*6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

sans objet

**XIII. DATA\_FAB\_PRODUICTS : DONNEES D'ITINERAIRES TECHNOLOGIQUES  
PLATEFORME LAIT**

**1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES**

*1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?*

Le jeu de données comprend les informations et mesures réalisées au cours des essais de fabrication de produits alimentaires mixtes à base de lait et d'un ingrédient végétal, en salle alimentaire ou sur la plateforme technologique de STLO, ou chez un partenaire du projet susceptible d'assurer ces fabrications dans des conditions alimentaires.

A priori, les données comprennent notamment

- les masses des différents intrants + produits et déchets émis
- les températures instantanées par intervalles
- les pH instantanés par intervalles
- les puissances absorbées instantanées par intervalles
- les débits instantanés de flux (dont eau) par intervalles
- et/ou les consommations cumulées d'énergies
- et/ou les consommations cumulées d'eau

Dans un souci de cohérence, les données de suivi de fermentation propres aux fabrications alimentaires sont incluses. Le produit de recherche DATA\_FERMENTATION sera alors plutôt dédié à la sélection de levains ou souches.

Des données issues de projets antérieurs, constituant les CONNAISSANCES PROPRES des partenaires telles que citées ou à citer dans l'accord de consortium en cours de rédaction, pourront également être mobilisées.

*1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?*

Les données seront :

- des images, typiquement des photographies, par exemple en format jpg ou png
- des tableaux, typiquement des séries temporelles ou des données croisées, en format tableur (par exemple xls, ods ou csv)
- du texte, typiquement des demandes et des rapports d'essai ou d'analyse, en format ouvert ou non, par exemple doc, odt ou pdf
- des présentations, typiquement pour de la transmission d'information ou la conceptualisation du workflow, en format ouvert ou non, par exemple ppt, odp ou pdf.

Le volume est estimé à 2 Go par copie du jeu de données.

## 2. DOCUMENTATION ET QUALITE DES DONNEES

### 2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?

Pour les produits qui seront réalisés sur la Plateforme Lait du STLO, le système de management de la qualité propre à ce partenaire sera utilisé. Il met en œuvre numéro d'action 23-010 propre au projet VEG&LAIT et la liasse documentaire propre à la demande de fabrication (fiche projet et fiche essai) et d'analyses (fiche de résultats d'analyses de composition). D'autres documents sont joints, tels que les résultats des analyses sanitaires émis par le laboratoire indépendant.

Les métadonnées à retenir concernent principalement le numéro du projet (23-010), le numéro d'essai, les noms des personnes impliquées dans l'essai (avec au moins 1 personne titulaire), la méthodologie appliquée et la date d'obtention de la donnée et les éventuelles versions (pooling, post-traitement). Dans l'encours du projet, ces métadonnées seront consignées dans le cahier de laboratoire et dans les fiches de liaisons entre services (par exemple formulaires de demande d'analyse).

STLO s'est doté d'un système de management de la qualité et la plupart de ses modes opératoires ont fait l'objet de la rédaction d'un protocole validé et stocké dans une base de données interne (LabCollector).

En fin de projet, ces métadonnées seront plus précisément définies dans la collection "VEG&LAIT" rattachée au dataverse de STLO sur Data INRAE.

### 2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?

Lors des essais expérimentaux, des points d'étape sont prévus tels que la calibration des appareils, l'inclusion de témoins ou références, la répétition des fabrications, le prélèvement d'un nombre représentatif d'échantillons, et la mise en œuvre de tests statistiques lors de l'exploitation des données.

La qualité des données sera également appréciée à dire d'experts, au cours d'une revue par les pairs (par exemple, comité de pilotage impliquant des personnes ayant l'expérience de ce type de fabrications).

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

## 3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE

### 3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?

Le système de management de la qualité de la Plateforme Lait permet le stockage, pendant 3 mois, des données mesurées automatiquement sur chaque procédé réservé et donc mobilisé pour chaque fabrication (séries temporelles stockées sur la base PCVue). Si ces données sont considérées comme nécessaires pour les objectifs du projet, le système permet aussi la création d'un entrepôt de données dans lequel sont téléversées automatiquement les informations des procédés mobilisés pour chaque essai (un entrepôt par essai). Articulées avec l'installation de compteurs, ces entrepôts pourront par exemple stocker des données cumulées de consommations d'énergies ou d'eau.

Il sera fait en outre, pendant l'en-cours de la tâche, au moins 3 copies :

- une copie locale sur le disque dur ou un HD externe de Fanny Guyomarc'h, Gilles Garric ou Gabrielle Jagorel (CDD sous la responsabilité de F. Guyomarc'h) , sans cryptage particulier.
- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

### **3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?**

Les données sensibles identifiées sont les données des procédés technologiques confidentielles. La gestion de la Plateforme Lait permet de restreindre l'accès aux zones d'équipements utilisés pour le projet, pendant la durée des essais. Les données issues des essais sont des RESULTATS au sens de l'accord de consortium et leur partage ou publication sont réglés par cet accord.

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE. Les entrepôts de données de la Plateforme STLO sont sécurisés et stockés pendant 3 ans.

## **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

### **4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?**

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

### **4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?**

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

### **4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?**

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

## **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

### **5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?**

Le WP2 s'échelonne de 2024 à 2027.

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

Les publications scientifiques et techniques, ainsi que les rapports de stage issus du projet seront dans la mesure du possible déposés dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

### **5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?**

Les données à conserver sont

- les fiches techniques des ingrédients utilisés
- les tableaux consommations cumulées de chaque intrant et de production cumulées de chaque coproduit ou déchet, à des fins d'inventaire du cycle de vie
- certaines séries temporelles, par exemple pH et température
- les tableaux et fichiers de résultats d'analyses physico-chimiques et sanitaires
- les métadonnées (fiche projet et fiche essai)

Une copie sera conservée sur un disque personnel (Fanny Guyomarc'h et/ou ingénieur-e contractuel-le recruté-e pour le projet), une autre sur l'espace personnel de Fanny Guyomarc'h alloué sur le datacenter, et une enfin sur l'espace NextCloud dédié du projet. Si les données le nécessitent, une copie peut être conservée 3 ans sur un entrepôt dédié attribué par la Plateforme Lait.

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE.

### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

La plupart des travaux du WP2 se tiendront à UMR STLO, qui utilise LabCollector pour le partage et l'archivage des méthodes expérimentales et de leurs révisions. Dans l'en-cours du projet, les méthodes et les réalisations seront également enregistrées et datées dans un cahier de laboratoire et un classeur physiques (papier) à défaut de numérisation. Dans l'exécution des manip, les logiciels propriétaires des machines seront utilisés in situ, mais comportent tous la possibilité d'exporter les données tabulées (exemple .csv) utilisables avec Excel. Une suite bureautique classique suffit.

### *5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

La publication des données est un objectif a priori, sans préjudice des termes de l'accord de consortium. Elle pourra être faite par un simple dépôt sur Data INRAE, et/ou par la production d'une communication scientifique exploitant les données (article de recherche et/ou data paper).

## **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

### *6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Fanny Guyomarc'h en tant que responsable du WP2.

### *6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

La curation et la mise à disposition sur Data INRAE est estimée à 1 mois-personne.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne.

## **XIV. DATA\_PROP\_PRODUIITS : DONNEES ANALYTIQUES DES COMPOSITIONS ET PROPRIETES FONCTIONNELLES DES PRODUITS**

### **1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES**

#### *1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?*

Le produit de recherche concerne la caractérisation de la composition et des propriétés fonctionnelles des produit alimentaires, commerciaux ou autoproduits (OBJ\_PRODUIITS). Les données sont typiquement obtenues par la mesure et l'analyse quantitative des réponses suivantes :



- composition en nutriments (extrait sec, protéines, lipides, amidon, cendres, fibres)
- profil granulométrique, masse volumique, couleur des produits
- aspect (homogénéité, brillance, couleur, odeur) en fonction de la recette et du procédé de fabrication, notamment : présence ou non d'une homogénéisation, d'une aromatisation, d'une cuisson ou traitement thermique.
- texture (viscosité, résistance à la pénétration ou dureté, élasticité, seuil de cassure ou d'écoulement, adhésivité)
- stabilité des produits dans le temps (observation de synérèse ou de décantation)

Des données préexistantes pourront être réutilisées, telles que les données fournisseur ou les données issues du projet LEGALIM partagé par VALOREX et STLO depuis 2020.

### *1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?*

Les données sont principalement tabulées et au format .csv ou .xls ou .odt. Des données aux formats image ou vidéo pourront compléter le jeu de données, les formats sont .pdf ou .png ou .jpg ou .avi ou .mov

Le volume est estimé a priori à environ 2 Go par copie du jeu de données.

## **2. DOCUMENTATION ET QUALITE DES DONNEES**

### *2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?*

Sauf dans le cas des analyses commandées en prestation, la réalisation de ces analyses est une expertise de l'unité STLO qui en détient le savoir-faire.

STLO s'est doté d'un système de management de la qualité et la plupart de ses modes opératoires ont fait l'objet de la rédaction d'un protocole validé et stocké dans une base de données interne (LabCollector).

Les métadonnées qui accompagneront les données seront typiquement le ou les auteur-trice-s, le rattachement au présent projet par le numéro du projet (23-010), le titre ou numéro de la collecte, la date de collecte et de ses éventuelles versions (pooling, post-traitement). Dans l'encours du projet, ces métadonnées seront consignées dans le cahier de laboratoire et dans les fiches de liaisons entre services (par exemple formulaires de demande d'analyse). En fin de projet, ces métadonnées seront plus précisément définies dans la collection "VEG&LAIT" rattachée au dataverse de STLO sur Data INRAE.

### *2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?*

L'analyse des compositions, l'analyse des propriétés physiques ou fonctionnelles des produits font l'objet de fiches Assurance Qualité en cours de rédaction au STLO. L'analyse de l'extrait sec ou de certains macronutriments, tels que les protéines, sont adossées à des normes ISO.

Lors des protocoles d'observations, d'analyses ou de mesures, des points d'étape sont prévus tels que le calibrage des instruments, l'inclusion de témoins ou références, la répétition des tests, le prélèvement d'un nombre représentatif d'échantillons, et la mise en œuvre de tests statistiques lors de l'exploitation des données.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

### 3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE

#### 3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?

Le système de management de la qualité de la Plateforme Lait permet le stockage, pendant 3 mois, des données mesurées automatiquement sur chaque procédé réservé et donc mobilisé pour chaque fabrication (séries temporelles stockées sur la base PCVue). Si ces données sont considérées comme nécessaires pour les objectifs du projet, le système permet aussi la création d'un entrepôt de données dans lequel sont téléversées automatiquement les informations des procédés mobilisés pour chaque essai (un entrepôt par essai). Articulées avec la demande d'analyse, ce système peut également entreposer des données d'analyses physico-chimiques sur les produits.

Par ailleurs, les données d'analyses peuvent être stockées indépendamment de la Plateforme Lait. Il en sera fait, pendant l'en-cours de la tâche, au moins 3 copies :

- une copie locale sur le disque dur ou un HD externe de Fanny Guyomarc'h, Gilles Garric ou Gabrielle Jagorel (CDD sous la responsabilité de F. Guyomarc'h) , sans cryptage particulier.
- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

#### 3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?

Les données sensibles identifiées sont les identités des membres des participants et participantes au projet VEG&LAIT, des volontaires recruté-es pour les entretiens dans les WP1 et WP4, et des volontaires recruté-es pour les tests consommateur-trices dans les WP2 et WP3 (voir section 4a) et les données couvertes par la propriété intellectuelle (voir section 4b).

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE. Les entrepôts de données de la Plateforme STLO sont sécurisés et stockés pendant 3 ans.

### 4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE

#### 4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITÉES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### 4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### 4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

## 5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME

### 5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?

Le WP2 s'échelonne de 2024 à 2027.

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

Les publications scientifiques et techniques, ainsi que les rapports de stage issus du projet seront dans la mesure du possible déposés dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

### 5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL Institut Agro.

### 5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?

La plupart des travaux du WP2 se tiendront à UMR STLO, qui utilise LabCollector pour le partage et l'archivage des méthodes expérimentales et de leurs révisions. Dans l'en-cours du projet, les méthodes et les réalisations seront également enregistrées et datées dans un cahier de laboratoire et un classeur physiques (papier) à défaut de numérisation. Dans l'exécution des manip, les logiciels propriétaires des machines seront utilisés in situ, mais comportent tous la possibilité d'exporter les données tabulées (exemple .csv) utilisables avec Excel. Une suite bureautique classique suffit.

### 5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

## 6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES

### 6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?

Fanny Guyomarc'h en tant que responsable du WP2.

### 6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?

La curation et la mise à disposition sur Data INRAE est estimée à 1 mois-personne.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne.

## XV. DATA\_CONFOCAL : IMAGES DE MICROSTRUCTURES DE PRODUITS

### 1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

#### 1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?

Le produit de recherche concerne la caractérisation de la microstructure des produit alimentaires, commerciaux ou autoproduits (OBJ\_PRODUIITS). Les données sont typiquement obtenues par imagerie confocale avec marquage spécifique des macronutriments, puis traitement d'image par exemple pour déterminer des colocalisations de constituants ou la porosité d'une structure.

Des données préexistantes pourront être réutilisées, telles que les données fournisseur ou les données issues du projet LEGALIM partagé par VALOREX et STLO depuis 2020.

#### 1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?

Les données seront en premier lieu des images au format .czi, qui seront aussi converties au format .tiff et/ou .jpg Un échantillon est caractérisé d'abord par plusieurs images correspondant à plusieurs grossissements (mode en 2D), puis éventuellement par images en mode 3D, constituées de stacks de plusieurs. Ceci générera des fichiers .czi de taille allant de 2 à 3 Go par échantillon en fonction du mode d'acquisition (2D vs 3D) et de la résolution. En complément de ces images, certaines données seront quantifiées (ex : porosité) par analyse d'image externe.

Le volume est estimé à 3 Go par échantillon, soit 30 à 50 Go par campagne et jusqu'à 500 Go par copie du jeu de données au terme du projet.

### 2. DOCUMENTATION ET QUALITE DES DONNEES

#### 2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?

Sauf dans le cas des analyses commandées en prestation, la réalisation de ces analyses est une expertise de l'unité STLO qui en détient le savoir-faire. STLO s'est doté d'un système de management de la qualité et la plupart de ses modes opératoires ont fait l'objet de la rédaction d'un protocole validé et stocké dans une base de données interne (LabCollector).

Les métadonnées qui accompagneront les données seront typiquement le ou les auteur-trice-s, le rattachement au présent projet par le numéro du projet (23-010), le titre ou numéro de la collecte, la date de collecte et de ses éventuelles versions (pooling, post-traitement). Dans l'encours du projet, ces métadonnées seront consignées dans le cahier de laboratoire et dans les fiches de liaisons entre services (par exemple formulaires de demande d'analyse). En fin de projet, ces métadonnées seront plus précisément définies dans la collection "VEG&LAIT" rattachée au dataverse de STLO sur Data INRAE.

#### 2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?

L'analyse des microstructures par microscopie fait l'objet de fiches Assurance Qualité en cours de rédaction au STLO. L'analyse de l'extrait sec ou de certains macronutriments, tels que les protéines, sont adossées à des normes ISO.

Lors des protocoles d'observations, des points d'étape sont prévus tels que le calibrage des instruments, l'inclusion de témoins ou références, la répétition des tests, l'observation d'un nombre représentatif d'échantillons, et la mise en oeuvre de tests statistiques lors de l'exploitation des données.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### **3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?**

Pendant l'en-cours de la tâche, il sera fait au moins 3 copies :

- une copie locale sur le disque dur ou un HD externe de Fanny Guyomarc'h, Jordane Ossemond et/ou du microscope confocal, sans cryptage particulier.
- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

#### **3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?**

Les données sensibles identifiées sont les données couvertes par la propriété intellectuelle (voir section 4b). Pour ce produit de recherche, il n'y a pas de risque identifié : sans connaissance des autres données telles que DATA\_FAB\_PRODUIITS, il ne peut être fait usage de DATA\_CONFOCAL.

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE. Les entrepôts de données de la Plateforme STLO sont sécurisés et stockés pendant 3 ans.

### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

#### **4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?**

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### **4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?**

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### **4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?**

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5Bfr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5Bfr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

### **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

#### **5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?**

Le WP2 s'échelonne de 2024 à 2027.

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

Les publications scientifiques et techniques, ainsi que les rapports de stage issus du projet seront dans la mesure du possible déposés dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

#### *5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL Institut Agro.

#### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

La plupart des travaux du WP2 se tiendront à UMR STLO, qui utilise LabCollector pour le partage et l'archivage des méthodes expérimentales et de leurs révisions. Dans l'en-cours du projet, les méthodes et les réalisations seront également enregistrées et datées dans un cahier de laboratoire et un classeur physiques (papier) à défaut de numérisation. Dans l'exécution des manip, le logiciel ZEN BLUE, propriété de ZEISS, sera utilisé in situ, mais comporte la possibilité d'exporter les données images sous un format .tiff ou .jpg utilisable avec un logiciel libre de lecture ou d'analyse d'image (ex. ImageJ).

#### *5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

## **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

#### *6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Fanny Guyomarc'h en tant que responsable du WP2.

#### *6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. La tâche est programmée de 2024 à 2026 incluses (cf candidature).

La curation et la mise à disposition sur Data INRAE est estimée à 1 mois-personne.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne.



## XVI. DATA\_PROTEOLYSE : DONNEES ANALYTIQUES DE DIGESTION, PROTEOLYSE OU COMPOSITIONS EN ACIDES AMINES

### 1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

#### 1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?

Lors de fabrication de produits mixtes, de nouvelles données de digestion statique seront produites qui auront pour but d'évaluer la digestibilité et le profil en acides aminés essentiels bioaccessibles dans les produits mixtes. Tous les outils sont disponibles au sein de l'équipe Bioactivité et Nutrition de STLO.

i) Fractionnement des produits avant ou après digestion, par précipitation au TCA, pour séparer les acides aminés libres + les peptides solubles (fraction bioaccessible), d'une part; et les peptides et protéines en suspension, d'autre part (fraction non accessible).

ii) Dosage de groupements NH<sub>2</sub> libres - méthode OPA : Les groupements aminés libres seront dosés dans des assemblages moléculaires ou matrices alimentaires soumis à des digestions in vitro, et à différents temps au cours de ces digestions, permettant d'évaluer ainsi la cinétique de protéolyse. La méthode est basée sur la réaction de l'O-phthaldialdéhyde (OPA) et du DTT avec les amines primaires pour former un composé détectable à 340 nm. La détection est faite par mesure de l'absorbance.

iii) Profils SDS-PAGE : Les protéines présentes dans les assemblages moléculaires ou matrices alimentaires soumis à des digestions in vitro, et à différents temps au cours de ces digestions, seront séparées et semi-quantifiées par électrophorèse sur gel d'acrylamide.

iv) Hydrolyse totale puis analyse quantitative des acides aminés par chromatographie liquide HPLC pour déterminer la quantité de chaque espèce dans les différentes fractions et calculer un indice de performance nutritionnelle de type PD-CAAS.

Là où c'est possible, les données d'analyses issues des fabrications du projet doctoral DeciBELL et du projet IMAGINABLE seront réutilisées. Les données de IMAGINABLE, en particulier, sont détenues par STLO et sont totalement disponibles.

#### 1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?

OPA : Données d'absorbance traitées dans un tableur : résultats exprimés en mg ou mmol par l de digestat ou par g de matière azotée. Les données numériques sont typiquement consignées dans un fichier du type .xls de taille de quelques centaines de ko.

SDS-PAGE : Les données seront en premier lieu des images de gels d'électrophorèse au format constructeur .scn, qui seront aussi converties au format .tiff et/ou .jpg. Chaque fichier .scn a une taille de quelques Mo. En complément de ces images, l'intensité des bandes d'électrophorèse sera quantifiée. Ces données numériques seront compilées dans des fichiers de type .xls ou .csv. La taille de ces fichiers est typiquement de l'ordre de quelques Mo.

HPLC : Données spectrales d'absorbance traitées dans un tableur : résultats exprimés en mg ou mmol par l de digestat ou par g de matière azotée. Les données numériques sont typiquement consignées dans un fichier du type .xls de taille de quelques centaines de ko.

Au total, on estime le volume de données à 1 Go par copie.

### 2. DOCUMENTATION ET QUALITE DES DONNEES

#### 2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?

Sauf dans le cas des analyses commandées en prestation, la réalisation de ces analyses est une expertise de l'unité STLO qui en détient le savoir-faire. La méthodologie a fait l'objet de publications antérieures (Ménard et al., 2018, 2018), y compris d'un protocole issu d'un consensus international (Brodkorb et al. 2019) auquel STLO a activement participé.

STLO s'est doté d'un système de management de la qualité et la plupart de ses modes opératoires ont fait l'objet de la rédaction d'un protocole validé et stocké dans une base de données interne.

Les métadonnées qui accompagneront les données seront typiquement le ou les auteur-trice-s, le rattachement au présent projet par le numéro du projet (23-010), le titre ou numéro de la collecte, les paramètres de la digestion (exécution du mode opératoire : quantité d'enzymes, température, agitation, prélèvements, quantité de pefabloc, etc), la date de collecte et de ses éventuelles versions (pooling, post-traitement). Dans l'encours du projet, ces métadonnées seront consignées dans le cahier de laboratoire et dans les fiches de liaisons entre services (par exemple formulaires

de demande d'analyse). En fin de projet, ces métadonnées seront plus précisément définies dans la collection "VEG&LAIT" rattachée au dataverse de STLO sur Data INRAE.

## **2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?**

La digestion statique *in vitro*, le dosage OPA (y-compris le traitement des données), l'analyse des protéines par SDS-PAGE et l'analyse des acides aminés par HPLC font l'objet de fiches Assurance Qualité en cours de rédaction au STLO.

Lors des protocoles expérimentaux, des points d'étape sont prévus tels que la calibration des appareils, l'analyse de témoins ou d'étalons, la répétition des productions, la répétition des analyses et la mise en oeuvre de tests statistiques lors de l'exploitation des données. Dans le cas de la digestion statique, des validations inter-laboratoires ont également été mises en oeuvre et publiées.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

## **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

### **3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?**

Pendant l'en-cours de la tâche, il sera fait au moins 3 copies :

- une copie locale sur le disque dur ou un HD externe de Valérie Gagnaire ou Fanny Guyomarc'h, sans cryptage particulier.
  - une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
  - une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet
- Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

### **3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?**

Les données sensibles identifiées sont les données couvertes par la propriété intellectuelle (voir section 4b). Pour ce produit de recherche, il n'y a pas de risque identifié : sans connaissance des autres données telles que DATA\_FAB\_PRODUIITS, il ne peut être fait usage de DATA\_PROTEOLYSE.

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE. Les entrepôts de données de la Plateforme STLO sont sécurisés et stockés pendant 3 ans.

## **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

### **4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?**

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont

accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagés par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

### **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

#### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Le WP2 s'échelonne de 2024 à 2027.

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

Les publications scientifiques et techniques, ainsi que les rapports de stage issus du projet seront dans la mesure du possible déposés dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

#### *5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL Institut Agro.

#### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

La plupart des travaux du WP2 se tiendront à UMR STLO, qui utilise LabCollector pour le partage et l'archivage des méthodes expérimentales et de leurs révisions. Dans l'en-cours du projet, les méthodes et les réalisations seront également enregistrées et datées dans un cahier de laboratoire et un classeur physiques (papier) à défaut de numérisation. Dans l'exécution des manips, les logiciels propriétaires des machines seront utilisés in situ, tels que le logiciel IMAGELAB de BIO-RAD, mais comportent tous la possibilité d'exporter les données tabulées (exemple .csv) utilisables avec Excel. Une suite bureautique classique suffit.

#### *5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

## 6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES

*6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Fanny Guyomarc'h et Valérie Gagnaire respectivement en tant que responsable du WP2 et responsable des analyses concernées.

*6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. La tâche est programmées de 2024 à 2026 incluses (cf candidature).

La curation et la mise à disposition sur Data INRAE est estimée à 0.5 mois-personne par jeu de données.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne par jeu de données.

## XVII. DATA\_SAFETY : DONNEES DE CONTRÔLE MICROBIOLOGIQUE DES FABRICATIONS

### 1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

*1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?*

Le produit de recherche concerne la recherche de microorganismes pathogènes ou d'altération dans les produit alimentaires, commerciaux ou autoproduits (OBJ\_PRODUIITS). Les données sont typiquement obtenues par dénombrements microbiologiques réalisés en sous-traitance par LABOCEA.

Des données préexistantes pourront être réutilisées, telles que les données fournisseur ou les données issues du projet LEGALIM partagé par VALOREX et STLO depuis 2020.

*1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?*

Question sans réponse.

### 2. DOCUMENTATION ET QUALITE DES DONNEES

*2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?*

Les analyses sanitaires sont réalisées par un laboratoire prestataire agréé et accrédité, qui détient ses méthodes de travail mais fournit des métadonnées spécifiques telles que la méthode de référence utilisée (ISO ou NF) et les preuves de son accréditation (COFRAC) et de ses agréments (Ministères). Les autres métadonnées qui accompagneront les données seront typiquement le ou les personnes ayant passé commande, le rattachement au présent projet par le numéro du projet (23-010), le titre ou numéro de la collecte, la date de collecte et de ses éventuelles versions (pooling, post-traitement). Dans l'encours du projet, ces métadonnées seront consignées dans le cahier de laboratoire et dans les fiches de liaisons entre services (par exemple formulaires de demande d'analyse). En fin de projet, ces métadonnées seront plus précisément définies dans la collection "VEG&LAIT" rattachée au dataverse de STLO sur Data INRAE.

*2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?*

L'accréditation et l'agrément du laboratoire prestataire font foi.

### 3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE

*3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Pendant l'en-cours de la tâche, il sera fait au moins 3 copies :

- une copie locale sur le disque dur ou un HD externe de Louis Delaunay et/ou Jeanne-Marie Membré, sans cryptage particulier.
- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

### **3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?**

Les données sensibles identifiées sont les données confidentielles des procédés technologiques, des résultats du plan de contrôle microbiologique réalisé soit sur la Plateforme Lait soit chez les partenaires, et les données d'enquête (dires d'experts). La gestion de la Plateforme Lait permet de restreindre l'accès aux zones d'équipements utilisés pour le projet, pendant la durée des essais. Les données issues de la tâche 5.4 sont des RESULTATS au sens de l'accord de consortium et leur partage ou publication sont réglés par cet accord.

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE. Les entrepôts de données de la Plateforme STLO sont sécurisés et stockés pendant 3 ans.

## **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

### **4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?**

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

### **4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?**

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagés par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

### **4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?**

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

## **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

### **5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?**

Le WP2 s'échelonne de 2024 à 2027.

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

Les publications scientifiques et techniques, ainsi que les rapports de stage issus du projet seront dans la mesure du possible déposés dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour

une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

*5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Ces données sont des points de contrôle et n'ont pas besoin d'être conservées sur le long terme.

*5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

Les rapports d'analyse livrés par les prestataires sont lisibles par un logiciel classique de lecture de document PDF, par ex. Acrobat.

*5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

Il est probable que la question soit sans objet pour ce produit de recherche, même si la donnée sera conservée par STLO.

## **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

*6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Fanny Guyomarc'h et Valérie Gagnaire respectivement en tant que responsable du WP2 et responsable des analyses concernées.

*6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

Question sans réponse.

## **XVIII. DATA\_FERMENTATIONS : DONNEES ANALYTIQUES DE SUIVI DES FERMENTATIONS**

### **1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES**

*1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?*

Les données qui seront produites dans le cadre du projet sont les suivantes :

- Données sur les identifiants numérique, taxonomique et d'origine des bactéries (Jeu de données)
- Données sur les paramètres de croissance des bactéries lactiques (Jeu de données)
- Données sur les métabolites primaires des bactéries lactiques : hydrolyse des sucres, production d'acides organiques, hydrolyse des protéines, indices de protéolyse (Jeu de données). Pour ce qui est relatif aux protéines, les données sont de même nature et traitées comme pour le produit de recherche DATA\_PROTEOLYSE qui englobe digestion et fermentation.

Des données préexistantes sont susceptibles d'être utilisées. Ce sont notamment :

- Données sur les identifiants des bactéries, livrés à l'achat de souches commerciales par des fabricants ou fournisseurs externes
- Données de croissance et de métabolisme issus de travaux antérieurs et conduits à STLO, tels que le projet doctoral DeciBELL ou les projets LETSPROSEED, LOCALNUTLEG ou PROBIOVEG.

*1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?*

Les données quantitatives générées seront aux formats .xls et .csv. Les données qualitatives seront au format image (jpeg ou tif).

Le volume des données sera inférieur à 1 Go, par copie du jeu de données.



## 2. DOCUMENTATION ET QUALITE DES DONNEES

### 2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?

Sauf dans le cas des analyses commandées en prestation, la réalisation de ces analyses est une expertise de l'unité STLO qui en détient le savoir-faire. La méthodologie a fait l'objet de publications antérieures, en particulier dans la thèse DeciBELL (Canon et al., 2020; 2021).

STLO s'est doté d'un système de management de la qualité et la plupart de ses modes opératoires ont fait l'objet de la rédaction d'un protocole validé et stocké dans une base de données interne (LabCollector).

Les métadonnées qui accompagneront les données seront typiquement le ou les auteur-trice-s, le rattachement au présent projet par le numéro du projet (23-010), le titre ou numéro de la collecte, la date de collecte et de ses éventuelles versions (pooling, post-traitement). Dans l'encours du projet, ces métadonnées seront consignées dans le cahier de laboratoire et dans les fiches de liaisons entre services (par exemple formulaires de demande d'analyse). En fin de projet, ces métadonnées seront plus précisément définies dans la collection "VEG&LAIT" rattachée au dataverse de STLO sur Data INRAE.

### 2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?

Le criblage de bactéries, l'observation de cinétiques de croissance et l'observation de réponses métaboliques ont fait l'objet de publications dans des revues à comité de lecture et/ou font l'objet de fiches Assurance Qualité en cours de rédaction au STLO.

Lors des protocoles expérimentaux, des points d'étape sont prévus tels que la calibration des appareils, l'analyse de témoins ou d'étalons, la répétition des analyses et la mise en oeuvre de tests statistiques lors de l'exploitation des données.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

## 3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE

### 3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?

Pendant l'en-cours de la tâche, il sera fait au moins 3 copies :

- une copie locale sur le disque dur ou un HD externe de Valérie Gagnaire, Gabrielle Jagorel (CDD sous la responsabilité de FG) et/ou Fanny Guyomarc'h, sans cryptage particulier.
  - une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
  - une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet
- Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

### 3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?

Les données sensibles identifiées sont les données couvertes par la propriété intellectuelle (voir section 4b).

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE. Les entrepôts de données de la Plateforme STLO sont sécurisés et stockés pendant 3 ans.

#### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

##### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

##### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagés par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

##### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

#### **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

##### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Le WP2 s'échelonne de 2024 à 2027.

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

Les publications scientifiques et techniques, ainsi que les rapports de stage issus du projet seront dans la mesure du possible déposés dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

##### *5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL AgroParisTech.

##### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

La plupart des travaux du WP2 se tiendront à UMR STLO, qui utilise LabCollector pour le partage et l'archivage des méthodes expérimentales et de leurs révisions. Dans l'en-cours du projet, les méthodes et les réalisations seront

également enregistrées et datées dans un cahier de laboratoire et un classeur physiques (papier) à défaut de numérisation. Dans l'exécution des manip, les logiciels propriétaires des machines seront utilisés in situ, mais comportent tous la possibilité d'exporter les données tabulées (exemple .csv) utilisables avec Excel. Une suite bureautique classique suffit.

#### **5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?**

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

### **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

#### **6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?**

Fanny Guyomarc'h et Valérie Gagnaire respectivement en tant que responsable du WP2 et responsable des analyses concernées.

#### **6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?**

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. La tâche est programmées de 2024 à 2026 incluses (cf candidature).

La curation et la mise à disposition sur Data INRAE est estimée à 0.5 mois-personne par jeu de données.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne par jeu de données.

## **XIX. DATA\_ACV : DONNEES D'INVENTAIRES ET D'ANALYSE DU CYCLE DE VIE**

### **1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES**

#### **1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?**

Le produit de recherche concerne l'établissement de l'inventaire de cycle de vie (LCI) et l'analyse des impacts environnementaux (ACV) pour la fabrication des ingrédients végétaux issus de la tâche T2.1 et pour la fabrication in fine des produits alimentaires mixtes issus de la tâche T2.2. La collecte des données nécessaires aux LCI sera réalisée au sein de ces tâches. L'ACV elle-même est réalisée en tâche T5.3.

En première intention, une gamme de 6 à 8 ingrédients est considérée (type : farine, concentrat ou isolat; espèces : féverole, pois, lentille, lupin ou pois chiche - voir OBJ\_INGREDIENTS). Trois types de produits sont considérés : "yaourts", crème dessert ou crème glacée (voir OBJ\_PRODUIITS). Pour chaque type de produit, un itinéraire technologique est mis en oeuvre avec différents scénarios, avec les facteurs de variation suivants :

- pour les ingrédients : l'espèce, les conditions agronomiques de production, le paramétrage du décorticage et celui de la turboséparation;
- pour les produits alimentaires : composition de la recette, présence ou non d'une étape d'homogénéisation, valeur de certains paramètres techniques (température, durée...).

Les données nécessaires pour la constitution de chaque LCI sont :

- le choix d'un flux de référence; par défaut : 1 fabrication de bout en bout (pour chaque inventaire)
- les différentes opérations unitaires mobilisées et leur enchaînement (description du système) y compris les opérations générales de l'installation telles que production de fluides, nettoyage en place ou traitement de l'eau
- les masses, matériaux, durée d'amortissement et durée d'occupation des machines pour 1 fabrication
- les énergies employées et leurs consommations, opération par opération
- pour les OBJ\_INGREDIENTS, les différentes matières relatives à la graine, entrantes (récolte) et sortantes (refus de tri, pellicules, farine, son, concentré, amidon etc), et pour chaque flux : les masses, les compositions, la destination, et le cas échéant le prix de chaque matière sortante si elle n'est pas un déchet
- pour les OBJ\_PRODUIITS, les différentes matières relatives au produit, entrantes (ingrédients des recettes, ferments...) et sortantes (pertes, coproduits), et pour chaque flux : les masses, les compositions, la destination, et le cas échéant le prix de chaque matière sortante si elle n'est pas un déchet

- pour tous, les différentes autres matières entrantes (produits de nettoyage, produits désinfectants : fongicides, insecticides..., emballages, eau de pousse, eau de nettoyage, eau d'échange de chaleur...) et sortantes (déchets ou autres destinations) : origine, les masses, les compositions, la destination, et le cas échéant le prix de chaque matière sortante i elle n'est pas un déchet (par exemple recyclage des matériaux)
- Les éventuels transports : matériau et masse transportée, distance, modalité, énergie, ambiance contrôlée ou non, volume occupé/disponible
- Les éventuels stockages : matériau et masse stockée, durée, énergie, ambiance contrôlée ou non, volume occupé/disponible

Ces données seront prioritairement recueillies sur site, c'est-à-dire chez les partenaires STLO, VALOREX et INVITATION A LA FERME, en présentiel par Fanny Guyomarc'h et en accord avec les dispositions de l'accord de consortium. Des données complémentaires ou substitutives pourront être utilisées, telles que celles fournies en 2022 à STLO par Terres Inovia à l'occasion d'une prestation pour le projet SYALSA IMAGINABLE.

Les LCI et leurs variations (sensibilité) constituent les livrables D2.1.4 et D2.1.5 pour OBJ\_INGREDIENTS et les livrables D2.3.2 et D2.3.4 pour OBJ\_PRODUIITS.

Les impacts environnementaux seront calculés après la saisie des données de l'inventaire sur le logiciel Simapro, mis à disposition des agents de l'INRAE par la plateforme MEANS co-gérée par l'INRAE et le CIRAD. La méthode choisie sera le Product Environmental Footprint 3.0 qui permet de calculer les 17 impacts mid-points et les 3 impacts end-points couramment publiés par la base de données publique AGRIBALYSE.

Les LCI et ACV sont circonscrits a priori sur un périmètre "gate-to-gate" c'est-à-dire transformation seulement, en prenant comme matière première agricole des données issues de la base de données publique AGRIBALYSE 3.1. Cependant, compte-tenu de la contribution importante de l'amont agricole dans les impacts environnementaux des produits agro-alimentaires, il sera préférable d'inclure également des données de premier plan, issues des itinéraires techniques agricoles des productions de VEG-ET-LAIT, au cours des années 2025 et suivantes. Elles pourront être introduites dans l'inventaire par le biais de l'outil MEANS IN-OUT (INRAE-CIRAD). Les données à collecter incluent le type de sol, sa pente, la succession culturale, les différentes interventions (engrais, amendements, protection, irrigation...) etc pour les cultures, et les données de cheptel, d'alimentation animale et de gestion des effluents pour les ateliers animaux (à préciser avec la plateforme MEANS).

## **1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?**

Les données collectées seront typiquement :

- un fichier Excel de volume environ 1 Mo par ICV (soit environ 20 Mo au total)
- une liasse de fichiers PDF, par exemple notices de produits commerciaux (machines, substances, etc), typiquement pour un volume de 50 Mo

Les données produites seront typiquement :

- le diagramme des flux, typiquement sur Powerpoint et si possible sur un outil de Material Flow Analysis (non disponible au T0 du projet) pour un volume de 0.1 Mo par ICV soit environ 2 Mo au total
  - les ICV mis à disposition sur Data INRAE (volume inconnu au T0 du projet)
  - les ICV saisis dans SIMAPRO par le biais de la plateforme MEANS (volume inconnu)
  - les résultats des ACV, sous la forme d'un fichier Excel d'environ 1 Mo par ICV (soit environ 20 Mo au total)
- Le volume total estimé est donc d'environ 100 Mo par copie du jeu de données.

## **2. DOCUMENTATION ET QUALITE DES DONNEES**

### **2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?**

Les métadonnées à rechercher et à retenir concernent :

- la traçabilité de la production (par exemple, si 1 récolte est utilisée pour fabriquer 1 ingrédient cible)
- l'entreprise responsable, le lieu et la date de la fabrication, et toute donnée permettant de se référer à son propre outil de traçabilité (par exemple ERP).

### **2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?**

Lors de mesures, des points d'étape sont prévus tels que la calibration des appareils, l'analyse de témoins ou d'étalons, la répétition des productions, la répétition des analyses et la mise en oeuvre de tests statistiques lors de l'exploitation des données.

La qualité des données d'ICV est généralement auto-évaluée par le biais d'une matrice de PEDIGREE. Sa mise en oeuvre nécessite de connaître la représentativité de l'ICV pour des systèmes de même nature; par exemple, il sera intéressant de recueillir des données de répétitions si elles ont lieu, de savoir si la graine provient d'un territoire particulier ou peut être considérée comme "France", ou de comparer une fabrication "témoin" censée représenter le "business as usual" avec d'autres ICV disponibles ou avec des données globalisées (par exemple de rendements

typiques pour l'usine ou pour le processus). A ce stade, l'expertise du partenaire TERRES INOVIA et celle du prestataire LEGGO seront précieuses. Pour la production laitière, les données de référence Grand Ouest ou France sont disponible dans la littérature de l'IDELE ou à l'UMR SAS qui est prestataire de VEG-ET-LAIT et hébergeur de MEANS.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>). Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### *3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Pendant l'en-cours de la tâche, il sera fait 3 copies :

- une copie locale sur le disque dur ou un HD externe de Fanny Guyomarc'h, sans cryptage particulier.
- une copie sur son espace alloué sur le DataCenter INRAE
- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

#### *3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Les données sensibles identifiées sont celles fournies par les entreprises, telles que par exemple des données de consommation énergétique. La transmission des données nécessaires à l'ICV devra donc se conformer aux dispositions de l'accord de consortium.

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE.

### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

#### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

### **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

#### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Sous réserve des dispositions de l'accord de consortium, l'INRAE encourage la mise à disposition publique des ICV sur Data INRAE. La tâche 5.2 s'échelonne de 2024 à 2027.

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

Les publications scientifiques et techniques, ainsi que les rapports de stage issus du projet seront dans la mesure du possible déposés dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

#### *5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Les données à conserver sont les tableaux d'inventaire complets, le schéma du système et les tableaux des résultats d'évaluation des impacts environnementaux. .

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL AgroParisTech.

#### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

Une suite bureautique classique suffit.

#### *5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données (ICV ou ACV) sont la publication d'un data paper et/ou une communication scientifique.

### **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

#### *6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Fanny Guyomarc'h en tant que responsable de la tâche 5.3.



## *6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

La curation et la mise à disposition sur Data INRAE est estimée à 0.5 mois-personne.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne.

## **XX. DATA\_SENSORIEL : DONNEES D'EVALUATION SENSORIELLE DES PRODUITS**

### **1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES**

#### *1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?*

De nouvelles données seront recueillies à l'issue de tests d'analyse sensorielle conduits auprès de différents publics :

\* des groupes d'individus recrutés parmi les partenaires directs du projet, pour une évaluation entraînée des ingrédients et des aliments fabriqués, notamment à STLO, IAF et/ou VALOREX;

\* des panels plus larges réalisés sur des échantillons de personnes volontaires "naïves" recrutées a priori au sein du personnel des partenaires, au delà des individus directement impliqués dans le projet. Au démarrage du projet, il n'est pas prévu de commander des analyses sensorielles plus poussées, auprès par exemple de l'Institut Agro (Rennes ou Dijon).

Chaque test fait l'objet d'un plan de contrôle. Des analyses libératoires sont prévues sur les produits dégustés et réalisées par un laboratoire sous-traitant indépendant.

Ces questionnaires proposent typiquement aux participant-e-s de qualifier un produit en verbalisant son appréciation vis-à-vis du produits

(recueil de verbatims) ou en évaluant son adhésion (échelle d'accord/pas d'accord) à des énoncés concernant le produit (recueil de fréquences

de réponses pour chaque niveau et chaque énoncé).

Des données pré-existantes, recueillies pendant des projets antérieurs tels que DIET+, LEG'ALIM ou IMAGINABLE pourront être utilisées à des fins de comparaisons et de discussion de nos propres résultats.

#### *1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?*

Les données sont typiquement numériques (notes, fréquences) et/ou textuelles (verbatim) et sont collectées dans des formats de type tableur. Le volume des données générées est estimé à [préciser].

### **2. DOCUMENTATION ET QUALITE DES DONNEES**

#### *2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?*

Pour les produits qui seront réalisés sur la Plateforme Lait du STLO, le système de management de la qualité propre à ce partenaire sera utilisé. Il met en oeuvre le numéro d'action 23-010 propre au projet VEG&LAIT et la liasse documentaire propre à la demande de fabrication (fiche projet et fiche essai) et d'analyses (fiche de résultats d'analyses de composition). D'autres documents sont joints, tels que les résultats des analyses sanitaires émis par le laboratoire indépendant.

Les métadonnées à retenir concernent principalement le numéro du projet (23-010), le numéro d'essai, les noms des personnes qui mettent en oeuvre l'analyse sensorielle, la méthodologie appliquée et la date d'obtention de la donnée.

#### *2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?*

Lors des protocoles d'analyse sensorielle, des points d'étape sont prévus tels que l'inclusion de témoins, la répétition des tests, le recrutement d'un nombre suffisant de volontaires, la mise en place de conditions de dégustation propres à éviter les biais, et la mise en oeuvre de tests statistiques lors de l'exploitation des données.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### *3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Il sera fait au moins 3 copies :

- une copie locale sur le disque dur ou un HD externe de Fanny Guyomarc'h et/ou de l'ingénieur-e en charge de l'exécution du WP2, sans cryptage particulier.
- une copie sur son/leur espace alloué sur le DataCenter INRAE
- une copie sur l'espace partagé dédié au projet

#### *3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Les données sensibles identifiées sont les identités des volontaires recrutés pour les tests. S'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE.

### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

#### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTologiques RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

## **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Le WP2 s'échelonne de 2024 à 2027.

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

Les publications scientifiques et techniques, ainsi que les rapports de stage issus du projet seront dans la mesure du possible déposés dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

### *5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Les données à conserver sont les tableaux des résultats complets d'évaluation sensorielle. Une copie sera conservée sur un disque personnel, une autre sur l'espace alloué sur le datacenter, et une enfin sur l'espace NextCloud du projet. La publication des données est un objectif a priori, sans préjudice des termes de l'accord de consortium. Elle pourra être faite par un simple dépôt sur Data INRAE, et/ou par la production d'une communication scientifique exploitant les données (article de recherche et/ou data paper). Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL AgroParisTech.

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

Une suite bureautique classique suffit.

### *5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

Le dépôt sur Data INRAE est un objectif a priori.

## **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

### *6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Fanny Guyomarc'h en tant que responsable du WP2.

### *6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

La curation et la mise à disposition sur Data INRAE est estimée à 0.5 mois-personne.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne.

## **XXI. DATA\_CONSO : DONNEES DE TESTS CONSOMMATEURS (PROJECTIVE MAPPING ET TESTS LONGITUDINAUX)**

### **1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES**

#### *1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?*

Les données seront produites pendant le WP3 qui consiste à questionner des consommateurs et consommatrices sur leurs perceptions et attitudes vis-à-vis de produits mixtes, en comparaison avec des produits laitiers ou végétaux de même nature, avant et après une période d'expositions périodiques à des produits mixtes au cours desquelles les sujets seront interrogés sur leur appréciation des qualités organoleptiques des produits. Au cours de ce WP3, plusieurs jeux de données seront générés qui constitueront le produit de recherche DATA\_CONSO. Il s'agit :

- des données de recrutement des sujets sur la base de données PanelSens gérée par UMR CSGA, avant et après sélection en fonction de critères d'inclusion + anonymisation
- des données de consentement et d'information des sujets participant à l'étude
- du ou des questionnaires de type CATA (adhésion à des affirmations) ou SI-FCQ (identification à des motivations de choix alimentaire) ou des procédures et supports de tests de projective mapping (comparaison implicite) et d'évaluation sensorielle (avec ou sans information) conçus pour l'étude,
- de données quantitatives issues des réponses (fréquences, notes... collectées dans les questionnaires et tests), avant et après curation, puis analyse
- de données qualitatives, de type verbatims issus d'entretiens dirigés, de questions ouvertes aux questionnaires ou de focus groups, avant et après curation, puis analyse.

Des données préexistantes pourront être réutilisées, telles que des résultats issus du projet IMAGINABLE, co-conduit par CSGA et STLO entre 2020 et 2023, dont la propriété est intégralement publique et dont les données sont donc disponibles.

#### *1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?*

Les données sont principalement tabulées, au format .xls ou .csv. Les données textuelles sont au format .txt ou .doc ou .odt ou .pdf. Des données d'image ou de vidéo pourront compléter le jeu de données, les formats sont .pdf ou .png ou .jpg ou .avi ou .mov

Le volume est estimé a priori à environ 0.5 à 1 Go par copie du jeu de données.

### **2. DOCUMENTATION ET QUALITE DES DONNEES**

#### *2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?*

Chaque dossier de données sera accompagné d'un fichier texte ou d'un document physique (cahier de labo) indiquant les métadonnées : il s'agit à minima d'identifier le contexte d'acquisition des données (auteur(s), contributeurs-trices, méthode(s) utilisée(s) pour acquérir ou analyser la donnée, événements maîtrisés ou non susceptibles de modifier l'organisation des tests et/ou d'avoir des conséquences sur les résultats), les dates de collecte et de révision et les unités de mesure dans certains cas.

#### *2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?*

L'unité CSGA s'est engagée, à travers tout son personnel, à renforcer les démarches entreprises dans la mise en place et la gestion du système d'assurance de la qualité et à prendre toutes les dispositions nécessaires pour assurer la fiabilité des résultats quantifiables et la traçabilité de ses activités. La cellule qualité est animée par deux agents et comporte un représentant de chaque équipe, plate-forme et service. Lors des protocoles de tests engageant des consommateurs et consommatrices, des points d'étape sont prévus tels que l'inclusion de témoins, la répétition des tests, le recrutement d'un nombre suffisant de volontaires, la mise en place de conditions de recrutement et de dégustation propres à éviter les biais, et la mise en oeuvre de tests statistiques lors de l'exploitation des données.

Ces points de contrôles sont notamment exercés par le comité d'éthique auprès duquel l'UMR CSGA obtient la validation du plan des travaux, préalablement à chaque campagne d'expérimentation.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données

proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées

permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### *3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Pendant l'en-cours de la tâche, il sera fait au moins 3 copies :

- une copie locale sur le disque dur ou un HD externe de Gaëlle Arvisenet et/ou de la ou le post-doctorant à recruter, sans cryptage particulier.
- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE
- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

#### *3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Les données sensibles identifiées sont les identités des membres des participants et participantes au projet VEG&LAIT, des volontaires recruté-es pour les entretiens dans les WP1 et WP4, et des volontaires recruté-es pour les tests consommateur-trices dans les WP2 et WP3 (voir section 4a) et les données couvertes par la propriété intellectuelle (voir section 4b).

Pour le WP3 et les tests mobilisant des consommateurs et consommatrices, l'anonymisation est assurée dès la sélection des sujets sur PanelSens. Seul(e)s les agent(e)s habilité(e)s de l'UMR CSGA, sous couvert de la direction d'unité, peuvent accéder à l'outil d'anonymisation par le biais d'un accès sécurisé par un mot de passe. Toute la suite des expérimentations sont réalisées en manipulant uniquement des codes participants. Les données individuelles (âge, genre, etc) sont traitées mais ne sont jamais reliées aux données discrètes issues des tests. Enfin, les données reliant les codes participants et les sujets sont supprimées une semaine après la fin de l'expérimentation (pour les éventuels besoins d'un rappel) ainsi que toute source de données rattachée à l'identité des sujets (par exemple les fichiers audio sont supprimés après transcription).

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE. Les entrepôts de données de la Plateforme STLO sont sécurisés et stockés pendant 3 ans.

### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

#### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFR%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFR%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

### **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

#### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Le WP3 s'échelonne de 2026 à 2028.

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

Les publications scientifiques et techniques, ainsi que les rapports de stage issus du projet seront dans la mesure du possible déposés dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

#### *5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL INSTITUT AGRO.

#### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

L'UMR CSGA mobilise des outils propriétaires pour concevoir et mettre en oeuvre les tests. Néanmoins les données curées et analysées sont livrées disponibles à une utilisation par une suite bureautique classique.

#### *5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

### **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

#### *6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Gaëlle Arvisenet est responsable du WP3.



## *6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. Le WP3 est programmé de 2026 à 2028 incluses (cf candidature).

La curation et la mise à disposition sur Data INRAE est estimée à 1 mois-personne.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne.

## **XXII. ENQ\_MARCHE : DONNEES ET ANALYSE ETUDE DE MARCHE DES PRODUITS MIXTES**

### **1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES**

#### *1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?*

Le produit de recherche concerne les données créées ainsi que les rapports intermédiaires et le rapport final d'une étude de marché commandée par l'INRAE auprès d'un tiers prestataire. Deux lots sont prévus, attribués chacun à un prestataire : le lot 1 concerne le profilage du ou des publics susceptibles d'être intéressés par les produits mixtes; et le lot 2 concerne le repérage et l'analyse d'innovations mixtes déjà mises sur le marché, en France ou à l'étranger, sur le segment des produits "laitiers" ultra-frais. Les données créées seront issues d'entretiens téléphoniques ou présentiels semi-directifs et de questionnaires en ligne diffusées auprès du grand public. Des données issues de recherche bibliographique pourront aussi être mobilisées, par exemple par la consultation des brevets, de la presse professionnelle, de rapports d'expert(e)s diffusés sur internet ou de la presse académique.

#### *1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?*

En amont de l'étude de marché elle-même, les données sont constituées de 6 fichiers texte constituant la liasse de l'appel d'offres et d'un tableau collectant les éléments de premiers entretiens avec des entreprises. L'ensemble occupe un volume de 3 Mo.

La liasse des offres reçues et leur examen seront également constitués de fichiers texte et tableur, pour un volume estimé à 10 Mo.

Pour l'enquête elle-même, il pourra s'agir d'enregistrements audio et de leur retranscription sous la forme de fichiers texte, de questionnaires internet de type Limesurvey et de la compilation de leurs résultats sous la forme d'un fichier tableur, de copies PDF de documents de la littérature grise, et enfin de fichiers présentation ou texte (supports des rapports intermédiaires et final de l'étude de marché).

L'ensemble occupera un volume estimé à 20 Mo par copie du jeu de données.

### **2. DOCUMENTATION ET QUALITE DES DONNEES**

#### *2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?*

Les métadonnées à retenir concernent principalement l'auteur (l'entreprise prestataire et ses mandataires), la méthodologie appliquée et la date d'obtention de la donnée.

#### *2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?*

Il y aura un contrôle par les pairs, en l'espèce 3 partenaires ayant l'habitude de commander des études de marché.

Le contrôle portera notamment sur la robustesse de l'échantillonnage pour le lot 1 (nombre de répondant-es, profilage au plus près de la démographie française au sens de l'INSEE) et sur la robustesse et la profondeur des analyses situationnelles pour le lot 2 (nombre de repérages, représentativité sur le segment ultra-frais, représentativité européenne voire française, profondeur des informations de contexte stratégique obtenues pour expliquer la mise sur le marché, les attentes et les résultats du lancement produit).

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### *3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Pendant l'en-cours de la tâche, il sera fait au moins 3 copies :

- une copie locale sur le disque dur ou un HD externe de Fanny Guyomarc'h, sans cryptage particulier.
- une copie sur son espace disque alloué sur le DataCenter INRAE
- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

#### *3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Les partenaires sont responsables de la confidentialité des informations qui leur sont confiées, aux termes de l'accord de consortium, qui inclut également la protection des données à caractère personnel (règlement européen EU 2016/679 du 27 avril 2016).

Selon le cahier des clauses administratives particulières (CCAP) de l'appel d'offres émis par l'INRAE pour sous-traiter l'étude de marché, l'entreprise prestataire est réputée devoir respecter les mêmes obligations de confidentialité et réaliser sa prestation en conformité au même règlement européen.

### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

#### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagés par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5Bfr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5Bfr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

### **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

#### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Les données ou analyse d'entretiens sous leurs formes élaborées et anonymisées feront l'objet d'un dépôt et d'un partage via l'espace NextCloud dédié au projet.

Les données ou entretiens en format bruts seront conservés par l'entreprise sous-traitante et stockés par elle par ses propres moyens. Néanmoins, à des fins de mise à disposition publique des produits de recherche financés par une subvention de l'ANR, les données brutes quantitatives et qualitative anonymisées font partie des livrables de l'étude de marché. Un embargo d'un maximum de 2 ans est susceptible d'être exigé par l'un ou l'autre des entreprises sous-traitantes des lot 1 et 2, et accordé par l'INRAE et/ou par les partenaires de l'accord de consortium, afin de lui conserver l'avantage compétitif de ce travail.

Les communications scientifiques et techniques issues du projet respecteront autant que possible les engagements de l'INRAE et de l'ANR pour une Science Ouverte. A minima, les partenaires viseront un dépôt des manuscrits dans une archive ouverte, dans le respect des dispositions de l'accord de consortium ou des politiques éditoriales des éditeurs (embargo, format...).

#### **5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?**

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, les données seront dans la mesure du possible publiées dans des entrepôts pérennes (ex. Data INRAE). Les publications déposées sur le HAL INRAE seront libres d'accès.

#### **5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?**

Une suite bureautique classique suffit.

#### **5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?**

Le dépôt de données dans un entrepôt tel que Data INRAE génère l'attribution d'un DOI. La publication d'un data paper sera envisagée, sans préjudice des termes de l'accord de consortium.

### **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

#### **6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?**

Fanny Guyomarc'h en tant que responsable de la tâche 5.1.

#### **6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?**

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. La tâche est programmée de 2024 à 2025 (cf candidature) et est intégralement supportée par la subvention. L'INRAE étant le partenaire commanditaire et acheteur de la prestation, il a spécifié dans le CCTP l'obligation de fournir un livrable contenant les données brutes curées, dans une forme réutilisable.

Le dépôt des données dans Data INRAE est estimé à 0.5 mois-personne.

## **XXIII. ENQ\_SAFETY : DONNEES D'ENQUETE - BIBLIO RISQUES SANITAIRES**

### **1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES**

#### **1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?**

Ce produit de recherche concerne principalement la réutilisation de données préexistantes, sous la forme d'une analyse de la littérature académique, de la presse professionnelle et d'informations issues d'organismes compétents tels que l'ANSES, pour l'identification des dangers microbiologiques et (éventuellement) chimiques susceptibles de concerner une filière de coproduction et cotransformation de lait et de légumineuses à des fins d'alimentation humaine générale. Ces données préexistantes pourront être complétées par des données nouvelles, telles que dires d'expert-es spécialement interrogé-es pour le projet.

L'ensemble de ces données constitue une enquête biblio.

## *1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?*

Les données sont principalement textuelles et au format .txt ou .doc ou .odt ou .pdf. Des données image pourront compléter le jeu de données, les formats sont .pdf ou .png ou .jpg. Des données tabulées pourront être produites au format .xls ou .csv.

Le volume est estimé a priori à moins de 0.5 Go par copie du jeu de données.

## **2. DOCUMENTATION ET QUALITE DES DONNEES**

### *2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?*

Les données académiques publiées, ou les données issues de l'ANSES ou d'ouvrages de référence, possèdent déjà des métadonnées formalisées (auteur(s), éditeur(s), date, DOI, ISBN, etc).

Pour les recueils de dires d'experts ou les données de littérature "grise", les métadonnées seront collectées de manière détaillée et homogène, par exemple sur une base de donnée ouverte de type Zotero. A minima la date, l'auteur et son affiliation, la personne en charge de la collecte seront identifiées. Tout élément susceptible de situer et préciser la donnée sera apporté (par exemple lieu, type de produit, etc).

### *2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?*

Les données académiques publiées, ou les données issues de l'ANSES ou d'ouvrages de référence, ont fait la preuve de leur qualité par le biais de la revue éditoriale et/ou de la revue par les pairs. La redondance et la convergence des informations issues de sources différentes, mais pertinentes pour le cas d'étude, constitue un élément de la représentativité des données. Le croisement avec les informations du Plan de Surveillance des Oléoprotéagineux, mis en oeuvre par le partenaire Terres Inovia, ou de tout autre plan de surveillance sanitaire connu et pertinent, contribuera à renforcer la qualité des données.

## **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

### *3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Pendant l'en-cours de la tâche, il sera fait au moins 3 copies :

- une copie locale sur le disque dur ou un HD externe de Louis Delaunay et/ou Jeanne-Marie Membré, sans cryptage particulier.

- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE

- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

### *3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Les données sensibles identifiées sont les données confidentielles des procédés technologiques, des résultats du plan de contrôle microbiologique réalisé soit sur la Plateforme Lait soit chez les partenaires, et les données d'enquête (dires d'experts). La gestion de la Plateforme Lait permet de restreindre l'accès aux zones d'équipements utilisés pour le projet, pendant la durée des essais. Les données issues de la tâche 5.4 sont des RESULTATS au sens de l'accord de consortium et leur partage ou publication sont réglés par cet accord.

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE. Les entrepôts de données de la Plateforme STLO sont sécurisés et stockés pendant 3 ans.

## **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données sensibles identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux

recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur"). Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagées par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

### **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

#### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Sous réserve des dispositions de l'accord de consortium, l'INRAE encourage la mise à disposition publique des données sur Data INRAE. La tâche 5.4 s'échelonne de 2026 à 2028.

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

Les publications scientifiques et techniques, ainsi que les rapports de stage issus du projet seront dans la mesure du possible déposés dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

#### *5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL de Oniris.

#### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

Une suite bureautique suffit.

#### *5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

## 6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES

*6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Louis Delaunay et Jeanne-Marie Membré en tant que responsables de la tâche 5.4 du projet.

*6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. La tâche est programmée de 2026 à 2028 (cf candidature).

Le dépôt des données dans Data INRAE est estimé à 0.5 mois-personne.

## XXIV. DATA\_CHALLENGE\_TESTS : DONNEES ANALYTIQUES DE SUIVI DES CHALLENGE TESTS

### 1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

*1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?*

Pour les principaux dangers identifiés par ENQ\_SAFETY, une évaluation quantitative des risques pour le consommateur sera réalisée, et des suggestions de stratégies d'atténuation pour les contrôler seront fournies. L'évaluation quantitative sera basée sur des modèles probabilistes, y compris l'incertitude et la variabilité (techniques de Monte Carlo du 2e ordre) et l'analyse de sensibilité.

L'évaluation pourra être vérifiée et affinée expérimentalement par la mise en oeuvre de challenge tests sur des produits commerciaux ou prototypes.

*1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?*

Les données sont principalement tabulées et au format .csv ou .xls ou .odt. Des données contextuelles aux formats image ou texte pourront compléter le jeu de données, les formats sont .pdf ou .png ou .jpg ou .doc.

Le volume est estimé a priori à environ 2 Go par copie du jeu de données.

### 2. DOCUMENTATION ET QUALITE DES DONNEES

*2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?*

Sauf dans le cas des analyses commandées en prestation, la réalisation de ces analyses est une expertise de l'unité SECALIM qui en détient le savoir-faire (ex Poissant et al., 2023). SECALIM s'est doté d'un système de management de la qualité et la plupart de ses modes opératoires ont fait l'objet de la rédaction d'un protocole validé.

Les métadonnées qui accompagneront les données seront typiquement le ou les auteur-trice-s, le titre ou numéro de la collecte, la date de collecte et de ses éventuelles versions (pooling, post-traitement). Dans l'encours du projet, ces métadonnées seront consignées dans le cahier de laboratoire physique, dans les entrées/sorties des opérations de modélisation, et/ou dans les fiches de liaisons entre partenaires (par exemple compte-rendus). En fin de projet, ces métadonnées seront plus précisément définies dans la collection "VEG&LAIT" rattachée au dataverse de STLO sur Data INRAE.

*2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?*

Lors des protocoles de modélisation, des points d'étape sont prévus tels que l'analyse de la sensibilité des résultats aux variations des conditions observées sur pour le cas d'étude, telles que par exemple la recette, ou le barème de cuisson (s'il y a une gamme de valeurs plutôt qu'une donnée initiale précise) ou la/les populations cibles (qui pourront être précisées par ENQ\_MARCHE) ou la sensibilité des résultats aux éventuels proxys de modèles, s'il n'y a pas de modèle spécifique au cas d'étude. Ces contrôles nécessitent le va-et-vient avec ENQ\_SAFETY et ENQ\_MARCHE



pour décrire le cas d'étude le plus finement possible au départ de la modélisation. Enfin, la variabilité et/ou l'incertitude des paramètres du modèle sont prise en compte par une évaluation par sondage de type Monte-Carlo.

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### ***3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?***

Pendant l'en-cours de la tâche, il sera fait au moins 3 copies :

- une copie locale sur le disque dur ou un HD externe de Louis Delaunay et/ou Jeanne-Marie Membré , sans cryptage particulier.

- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE

- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

#### ***3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?***

Les données sensibles identifiées sont les données confidentielles des procédés technologiques, des résultats du plan de contrôle microbiologique réalisé soit sur la Plateforme Lait soit chez les partenaires, et les données d'enquête (dires d'experts). La gestion de la Plateforme Lait permet de restreindre l'accès aux zones d'équipements utilisés pour le projet, pendant la durée des essais. Les données issues de la tâche 5.4 sont des RESULTATS au sens de l'accord de consortium et leur partage ou publication sont réglés par cet accord.

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE. Les entrepôts de données de la Plateforme STLO sont sécurisés et stockés pendant 3 ans.

### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

#### ***4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?***

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagés par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

### **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

#### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Sous réserve des dispositions de l'accord de consortium, l'INRAE encourage la mise à disposition publique des données sur Data INRAE. La tâche 5.4 s'échelonne de 2026 à 2028.

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

Les publications scientifiques et techniques, ainsi que les rapports de stage issus du projet seront dans la mesure du possible déposés dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

#### *5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL de Oniris.

#### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

Le logiciel R est utilisé pour la modélisation mais peut être communiqué en format texte. Pour le reste des données, une suite bureautique suffit.

#### *5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

### **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

#### *6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Louis Delaunay et Jeanne-Marie Membré en tant que responsables de la tâche 5.4 du projet.

## *6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. La tâche est programmée de 2026 à 2028 (cf candidature).

La curation et la mise à disposition sur Data INRAE est estimée à 1 mois-personne.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne.

## **XXV. DATA\_HOLISTIC : DONNEES D'EVALUATION HOLISTIQUE / BALANCE BENEFICES-RISQUES**

### **1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES**

#### *1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?*

Ce produit de recherche concerne l'évaluation risques-bénéfices du ou des scénario(s) de production et transformation à la ferme de produits mixtes à base de lait et de légumineuses, proposés par les WP1 et 2.

La démarche consiste à bâtir une matrice à double entrée : d'une part les paramètres actionnables du/des scénarios (par exemple l'espèce cultivée ou la recette fabriquée) et d'autre part les indicateurs pertinents pour l'évaluation des performances des scénarios sur chaque dimension considérée (économique, environnementale, nutritionnelle, sensorielle, sanitaire).

Les données qui seront produites consistent donc en matrices originales, appropriées au cas d'étude. Elle consistent également en résultats de l'évaluation risques-bénéfices, soit bruts (vecteur ou profil ou radar de notes sur toutes les dimensions) soit agrégés (et requérant des pondérations décidées en consultant les partenaires du projet).

De nombreuses données préexistantes seront réutilisées, en particulier les données issues du projet, telles que ENQ\_MARCHE, ENQ\_SAFETY, ENQ\_IDEA4, DATA\_PROP-PRODUITS, DATA\_SENSORIEL, DATA\_CONSO ou DATA\_ACV.

#### *1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?*

Les données sont principalement tabulées et au format .csv ou .xls ou .odt. Des données contextuelles aux formats image ou texte pourront compléter le jeu de données, les formats sont .pdf ou .png ou .jpg ou .doc.

Le volume est estimé a priori à environ 2 Go par copie du jeu de données

### **2. DOCUMENTATION ET QUALITE DES DONNEES**

#### *2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?*

La réalisation de la tâche 5.5 de VEG&LAIT (analyse risques-bénéfice) mobilise une expertise de l'unité SECALIM qui en détient le savoir-faire (ex Annunçao et al., 2019 et travaux projet HOLiFOOD). SECALIM s'est doté d'un système de management de la qualité et la plupart de ses modes opératoires ont fait l'objet de la rédaction d'un protocole validé. Les métadonnées qui accompagneront les données seront typiquement le ou les auteur-trice-s, la date de réalisation et de ses éventuelles versions (pooling, post-traitement). Dans l'encours du projet, ces métadonnées seront consignées dans le cahier de laboratoire physique, dans les entrées/sorties des opérations de modélisation, et/ou dans les fiches de liaisons entre partenaires (par exemple compte-rendus). En fin de projet, ces métadonnées seront plus précisément définies dans la collection "VEG&LAIT" rattachée au dataverse de STLO sur Data INRAE.

#### *2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?*

Lors de l'évaluation risques-bénéfices, des points d'étape sont prévus tels que la participation dirigée des partenaires (experts des données qu'ils produisent) ainsi que l'analyse de la sensibilité des résultats aux variations des conditions observées sur pour le cas d'étude (par exemple incertitudes sur des rendements ou des impacts environnementaux).

Les membres INRAE et/ou Institut Agro sont invités à suivre les formations à la gestion de données proposées par leurs institutions (pour INRAE : <https://callisto-formation.fr/mod/page/view.php?id=14859>).

Par ailleurs, sur la plateforme NEXTCLOUD du Projet des documents de formation et information relatifs aux Plan de Gestion des données sont mis à disposition des équipes.

Chaque fichier de données devra être accompagné d'une fiche descriptive de métadonnées

permettant d'identifier à minima :

- L'intitulé des données
- La ou les personnes les ayant recueillies
- Le format des données
- La date d'acquisition des données
- Chaque évolution du jeu de données donnera lieu à une incrémentation du versionnement
- L'adresse de stockage des données : la plateforme NEXTCLOUD du projet VEG&LAIT et à terme l'entrepôt (dataverse) dans DATAGOUV.

Et toute autre information qui semblera nécessaire à la description des données.

L'équipe projet assure la sensibilisation au respect des procédures qualité du personnel temporaire et des nouveaux arrivants par des offres de formations régulièrement.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### *3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Pendant l'en-cours de la tâche, il sera fait au moins 3 copies :

- une copie locale sur le disque dur ou un HD externe de Louis Delaunay et/ou Jeanne-Marie Membré , sans cryptage particulier.

- une copie sur leurs espaces respectifs alloués sur le DataCenter INRAE

- une copie sur la plateforme NEXTCLOUD partagée et dédiée au projet

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

#### *3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?*

Les données sensibles identifiées sont les données confidentielles des procédés technologiques et les données d'enquête (dires d'experts) permettant de pondérer les différentes dimensions de l'évaluation multicritère. Les données issues de la tâche 5.5 sont des RESULTATS au sens de l'accord de consortium et leur partage ou publication sont réglés par cet accord.

Le DataCenter et l'espace NextCloud sont des outils sécurisés par les services informatiques de l'INRAE. Les entrepôts de données de la Plateforme STLO sont sécurisés et stockés pendant 3 ans.

### **4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE**

#### *4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITEES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?*

Les données à caractère personnel identifiées sont les identités des personnes collaborant au projet, des volontaires recrutés pour les entretiens dans les WP1, WP3 et WP4, des expert-es consulté-es dans le WP5, et des sujets des tests d'analyse sensorielle en WP2 ou d'études consommateur en WP3. Outre les dispositions énoncées en 2b, s'il s'avère nécessaire de conserver ces identités, un formulaire d'information et de consentement leur sera remis, conformément aux recommandations de l'INRAE (<https://donnees-personnelles.intranet.inrae.fr/>). Des modèles sont accessibles ici : <https://donnees-personnelles.intranet.inrae.fr/les-demarches-de-mise-en-conformite/aide-au-choix-des-formalites/fiches-pratiques-referentiel> (typiquement "référentiel compte utilisateur").

Le DataCenter et l'espace NextCloud, où les données sont susceptibles d'être stockées, sont des outils sécurisés par les services informatiques de l'INRAE.

#### *4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?*

Un accord de consortium est en cours d'écriture entre le service partenariat du Centre INRAE Bretagne-Normandie et les autres parties du projet VEG&LAIT, à savoir : VALOREX, INVITATION A LA FERME, TERRES INOVIA. Dans l'intervalle et ce depuis la préparation de la candidature auprès de l'ANR, les parties sont liées par un accord de confidentialité (Non Disclosure Agreement). Les prestataires engagés par le consortium ou par l'une ou l'autre de ses parties sont également engagés par contrat à la confidentialité des données qui leur sont confiées et des résultats qui sont produits.

#### *4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?*

L'INRAE et ses agent-es respectent une charte de déontologie ([https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie\\_INRAE%5BFr%5D\\_Print\\_0.pdf](https://www.inrae.fr/sites/default/files/pdf/Charte-Deontologie_INRAE%5BFr%5D_Print_0.pdf)). Un comité ad hoc peut être saisi en cas de nécessité.

### **5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME**

#### *5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?*

Sous réserve des dispositions de l'accord de consortium, l'INRAE encourage la mise à disposition publique des résultats sur Data INRAE. La tâche 5.4 s'échelonne de 2026 à 2028.

Lorsque le livrable est achevé, un versement dans la collection "VEG&LAIT" sera réalisé. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>.

Les publications scientifiques et techniques, ainsi que les rapports de stage issus du projet seront dans la mesure du possible déposés dans une archive ouverte (HAL) tout en respectant les recommandations de la Direction INRAE pour une science ouverte (DIPSO) encourageant l'apposition d'une licence CC-BY sur les oeuvres, et tout en respectant les politiques éditoriales des éditeurs (embargo).

#### *5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?*

Pour rendre les données librement accessibles et réutilisables pour d'autres travaux, en accord avec les principes FAIR, elles seront dans la mesure du possible déposées dans la collection "VEG&LAIT" de l'entrepôt de données DATA INRAE. Cette collection est rattachée au dataverse UMR-STLO : <https://entrepot.recherche.data.gouv.fr/dataverse/stlo>. Ce service est garanti 10 ans.

Les travaux seront publiés dans des revues nativement en open access ou dans des revues hybrides couvertes par l'accord Couperin, puis rendues libres d'accès sur le HAL INRAE et le HAL de Oniris.

#### *5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?*

La plateforme MEANS (INRAE-CIRAD) et par exemple son outil DEXi ou un outil équivalent sur un autre accès public, seront utilisés. L'output des résultats sera en format tabulé ou texte, pour lesquels une suite bureautique suffit.

#### *5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?*

Le dépôt sur Data INRAE est un objectif a priori. D'autres manières d'obtenir un DOI pour le jeu de données sont la publication d'un data paper et/ou une communication scientifique.

### **6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES**

#### *6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?*

Louis Delaunay et Jeanne-Marie Membré en tant que responsables de la tâche 5.4 du projet.

#### *6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?*

La production et la valorisation des données sont financées par le projet VEG&LAIT, et donc par une subvention ANR complétée par l'engagement des ressources des partenaires du projet. La tâche est programmée de 2026 à 2028 (cf candidature).

La curation et la mise à disposition sur Data INRAE est estimée à 1 mois-personne.

La préparation et la soumission d'un data paper est estimée à 1 mois-personne.