



HAL
open science

Plan de Gestion des Données (PGD) scientifiques de l'UMRH

Romain Lardy, Céline Bobby, Mylène Delosiere, Dominique Pomies, Milka Popova, Hieu Nguyen-Ba, Vincent Largeau, Luciano Barreto-Mendes, Jéméry Tournayre

► **To cite this version:**

Romain Lardy, Céline Bobby, Mylène Delosiere, Dominique Pomies, Milka Popova, et al.. Plan de Gestion des Données (PGD) scientifiques de l'UMRH. INRAE. 2024, pp.20. hal-04645317v1

HAL Id: hal-04645317

<https://hal.inrae.fr/hal-04645317v1>

Submitted on 11 Jul 2024 (v1), last revised 12 Jul 2024 (v2)

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

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



Distributed under a Creative Commons Attribution 4.0 International License

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<p>AQ1299 V1</p>

Plan de Gestion des Données (PGD) scientifiques de l'UMRH

<p>1-Date de Création V1 Juillet 2023</p>	<p>2-Auteurs V1 R. Lardy ; C. Boby; M. Delosiere; D. Pomies; M. Popova; H. Nguyen-Ba; V. Largeau; L. Barreto-Mendes ; J. Tournayre</p>	<p>3-V1 Revue par : Groupe Assurance Qualité et Direction UMRH</p>
<p>4-Date Nouvelle Version</p>	<p>5-Auteurs de la Nouvelle version</p>	<p>6-Nouvelle version Revue par :</p>
<p>Modifications apportées au document par rapport à la version précédente</p>		

DOI (version publiée du plan de gestion) :XXX

Sommaire (clic droit>mettre à jour les champs)

1. Objet.....	2
2. Domaine d'application.....	2
3. Documents de références.....	2
4. Définitions et/ou Principe.....	3
5. Liste de diffusion.....	3
6. Description du plan de gestion des données.....	3
6.1. Informations sur la structure.....	3
6.2. Présentation générale des données.....	4
6.3. Droits de propriété intellectuelle.....	7
6.4. Sensibilité des données / Confidentialité.....	7
6.5. Partage des données.....	11
6.6. Organisation et documentation des données.....	17
6.7. Stockage et sécurité des données.....	19

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<p>AQ1299 V1</p>

6.8. Archivage et conservation des données.....	21
Annexe 1 – Licences logicielles.....	23
1. Freeware ou « logiciel gratuit ».....	23
2. Licence libre.....	23
2.1. Licences libres sans copyleft.....	23
2.2. Licence Libre avec copyleft.....	23
2.2.1. Copyleft faible ou limité.....	23
2.2.2. Copyleft fort.....	23
3. Licence propriétaires.....	23

1. Objet

Cette procédure décrit le cycle de vie des données scientifiques gérées par l'ensemble des agents de l'unité et les pratiques de l'unité pour la sécurisation et l'ouverture des données.

2. Domaine d'application

Données scientifiques de l'UMR Herbivores

3. Documents de références

- <https://science-ouverte.inrae.fr/> dont la note INRAE « Principes pour la gouvernance des données, des codes et des algorithmes » du 2 septembre 2022
- <https://entrepot.recherche.data.gouv.fr/dataverse/inrae>
- <https://doranum.fr/>
- « OUVERTURE DES DONNÉES : ... AUSSI OUVERT QUE POSSIBLE ... AUSSI FERMÉ QUE NÉCESSAIRE » : <https://science-ouverte.inrae.fr/fr/les-donnees-et-le-numerique-scientifiques/la-gouvernance-des-donnees/logigramme-aussi-ouvert-que>
- FICHE Données de la recherche et science ouverte : le principe d'un accès « aussi ouvert que possible, aussi fermé que nécessaire » https://science-ouverte.inrae.fr/sites/default/files/2023-06/Fiche_SO_Juridique_ouverture_vf.pdf

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<h2>AQ1299</h2> <h3>V1</h3>

- Aide à la décision sur la diffusion des données de recherche : <https://www.loginos.net/base/ylxxVO>
- Choix des licences : https://science-ouverte.inrae.fr/sites/default/files/2023-06/20221114-Ouverture%2Bdes%2BCodes%2BAlgorithmes%2Bet%2BLogiciels_VF.pdf et <https://science-ouverte.inrae.fr/fr/les-donnees-et-le-numerique-scientifiques/partager-publier-des-donnees-et-des-codes/comment-choisir>

4. Définitions et/ou Principe

PGD : Plan de Gestion des Données

5. Liste de diffusion

Tous les agents UMRH titulaires et non permanents

6. Description du plan de gestion des données

Basé sur le modèle PGD structure INRAE v15 (juin 2022)

6.1. Informations sur la structure

Nom : UMR 1213 UMRH (Unité Mixte de Recherche sur les Herbivores)

Type de structure : Unité de recherche

Identifiant de la structure : <https://ror.org/O3yvemy54>

Responsabilités dans la structure :

- DU : A. Ferlay
- DUA Stratégie scientifique : G. Cantalapedra
- DUA Opérationnel : Y. Faulconnier
- DUA RH : C. Ginane
- Référent VetAgro Sup : C. Chassaing
- Responsable sécurité des données : J. Domagalski

Etablissements tutelles :

- INRAE

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<h2>AQ1299</h2> <h3>V1</h3>

- VetAgro Sup

Département de rattachement INRAE

- PHASE : Physiologie Animale et Systèmes d'Elevage
- ECOSOCIO : Economie et Sociologie pour l'Agriculture, l'Alimentation et l'Environnement

Financeurs (permettant l'acquisition des jeux de données – hors projet)

- Tous les financeurs de l'UMR

6.2. Présentation générale des données

6.2.1. Mode d'obtention des données :

- Produites par un tiers
- Générées par la structure

6.2.2. Origine des données:

- Analyse
- Agrégation
- Code
- Enquête
- Expérimentation
- Observation
- Simulation, modélisation

6.2.3. Type de données :

- Audiovisuel : Vidéo ; Enregistrement audio
- Jeux de données numériques : Base de données ; Données issues d'un logiciel spécifique (exemple Chromatogramme ; PCR quantitative ; Séquençage ; Enregistrement rythme cardiaque ; Données géographiques dont GPS)
- Image
- Modèle
- Service (dont webservice)

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<h2>AQ1299</h2> <h3>V1</h3>

- Logiciel
- Texte : Tabulé ; Séquence ; Textuelle ; Rapport d'analyse ; Code ; Données Hiérarchiques
- Workflow
- Autres : Binaires (exécutable de logiciel par exemple) ; Sauvegarde

6.2.4. Format des données

Le tableau ci-dessous présente la correspondance entre le type de données, leurs formats et les logiciels spécifiques nécessaires à leur lecture/traitement.

Type données	Format	Logiciel spécifique nécessaire
Tabulée	.xls, .xlsx, .ndj, .sql, .xad, .csv, .odf	Oui pour .xls et .xlsx
Image	.jpeg, .tiff, .png, .raw, .pdf, .pptx, .bmp	
Séquence	.fastq	
Chromatogramme	.cna, .cpf	Oui
PCR quantitative	.eds, .sds, .edt	Oui
Code	.R, .py, .php, .c, .cpp, .cs, .java, .Rmd, .Rb, .pl, .bash, .sas, .mat, .xlsm, .vba, .apk, .bat	Oui
Textuelle	.doc, .docx, .txt, .odt, .ppt, .pptx, archives papier, .pdf, .tex	Oui pour .docx et .doc
Base de données	.accdb, .sql	Oui pour .accdb
Rapport d'analyse	.Rmd, .html, .pdf	
Vidéo	.mp4, .mov, .avi, .mpeg, .wma, .mkv,	

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<p>AQ1299 V1</p>

	<i>.wmv, .asf</i>	
Enregistrement rythme cardiaque	<i>.mkt</i>	<i>Oui</i>
Hiérarchique	<i>.xml</i>	
Enregistrement audio	<i>.mp3, .wav, .pcm, .wma, .midi, .ogg, .flac, .acc</i>	
Données géographiques dont GPS	<i>.GPX, .SHP</i>	
Sauvegarde	<i>.cpt</i>	<i>Diapason (Idèle)</i>

Une partie des données issues des unités ou installations expérimentales INRAE sont directement issues des systèmes d'information gérés par le CATI (Centre Automatisé de Traitement de l'Information) SICPA (Systèmes d'Informations et Calcul pour le Phénotypage Animal) dont le PGD associé est disponible (<https://doi.org/10.57745/S19P9X>)

6.2.5. Périmètre thématique des données

- *Animal Breeding and Animal Products*
- *Farming Systems and Practices*
- *Rural and Agricultural Sociology*
- *Biodiversity and Ecology*
- *Animal Health and Pathology*
- *Economics*
- *Animal Nutrition*
- *Microorganisms*
- *Food Safety*
- *Food and food processing*
- *Plant Breeding and Plant Products*
- *Omics*

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<p>AQ1299 V1</p>

- Information management
- Computer science/statistics
- Biochemistry and biochemical engineering

6.3. Droits de propriété intellectuelle

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

Toutes les données générées au sein de l'UMRH sont la copropriété d'INRAE et de VetAgro Sup, comme mentionné dans l'accord-cadre. Dans le cas de projets pour lesquels un accord de consortium a été rédigé, cet accord prévaut.

[En attente reformulation par le service partenariat] Dans le cadre de prestations de service, la propriété intellectuelle est cédée au demandeur de la prestation, sauf si accord spécifique dans le contrat.

Au sein d'INRAE, la Direction Partenariat et Transfert pour l'Innovation (DPTI) et le service partenariat du centre Auvergne-Rhône-Alpes sont en charge des droits de propriété intellectuelle, du transfert et de la protection des données.

6.4. Sensibilité des données / Confidentialité

6.4.1. Identification du niveau de sensibilité des jeux de données

Le niveau de sensibilité des jeux de données est défini projet par projet, selon la nature du projet ou du contrat, sauf les données à caractère personnel qui sont confidentielles et doivent respecter le Règlement Général sur la Protection des Données (RGPD).

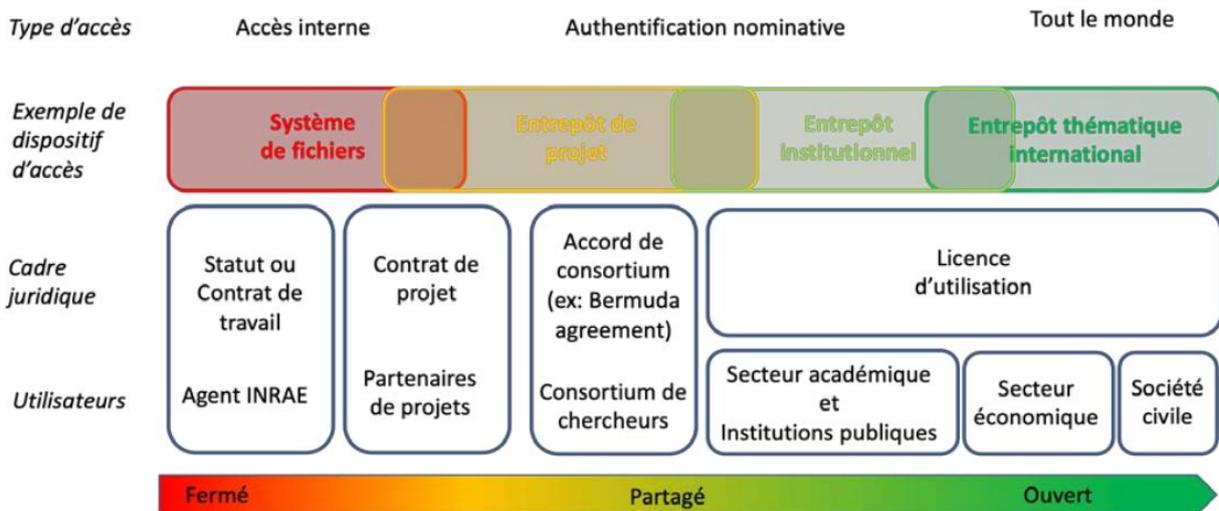
Confidentiel : données sensibles notamment à caractère personnel. Ces données sont fermées et à usage strictement interne à l'institut.

- **Diffusion restreinte** : données dont l'usage par des personnes externes à l'institut est possible sous certaines conditions. Cet usage peut être contrôlé via une authentification ou un dispositif qui permet d'identifier le ré-utilisateur. Cela concerne certains projets de recherche avec des partenaires privés ou présentant des contraintes notamment liées à la nature des données.
- **Diffusion limitée** : données "non achevées", par exemple lorsque l'exploitation scientifique, la publication des résultats ou la valorisation via la propriété

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<h2>AQ1299</h2> <h3>V1</h3>

intellectuelle n'est pas terminée, et dans le cas de certains projets de recherche avec des partenaires privés. Il est possible de n'ouvrir qu'une partie des données, ou de n'ouvrir qu'après traitement des données (bruitage, modification d'échelle, anonymisation ...), si tant est que cette information garde du sens sur le plan scientifique.

- **Publique** : toutes les autres données, dans le cadre du plan national pour la science ouverte. Ces données sont réutilisables par tous, gratuitement et sans aucun contrôle.



[Note INRAE « Principes pour la gouvernance des données, des codes et des algorithmes » du 2 septembre 2022]

BONNE PRATIQUE : Pour vous aider à décider si vous pouvez partager les données, vous pouvez utiliser cet outil du CIRAD : <https://www.loginos.net/base/ylxxVO>

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

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<h2>AQ1299</h2> <h3>V1</h3>

La sécurité des données sensibles est assurée par une formation des agents et un accès limité aux données avec une gestion des droits d'accès et un accès par authentification, ainsi qu'un partage limité :

- Ordinateur personnel : données liées au comptes utilisateurs (Mes documents), non accessibles aux autres utilisateurs
- NAS : groupes de sécurité
- NextCloud/Sharepoint : gestion des groupes et des droits d'accès
- Forges (code) : gestion des groupes et des droits d'accès

Les agents se conforment à la charte informatique (AQ1059, par exemple, il est interdit à INRAE d'installer sur son ordinateur un OneDrive, DropBox ou assimilés).

En cas d'accord de confidentialité, les thèses pourront être soutenues à huis clos.

Le chiffrement des données pourra être envisagé si nécessaire.

Les données d'enquêtes et les données administratives (Ressources Humaines notamment) contiennent des données personnelles qui doivent respecter le RGPD (cf paragraphe suivant).

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

Le responsable du traitement de données à caractère personnel est astreint à une obligation de sécurité. Il prend les mesures nécessaires pour garantir l'intégrité, la confidentialité et la disponibilité des données à caractère personnel, c'est à dire pour empêcher que les données ne soient déformées, endommagées ou que des personnes non autorisées y aient accès. Cette personne est définie avant la collecte des données (cf. informations RGPD).

Les données contenues dans les fichiers ne peuvent être consultées que par les personnes autorisées à y accéder en raison de leurs fonctions et/ou rôles dans ce traitement.

Les personnes concernées ont bénéficié d'une sensibilisation concernant la protection des données personnelles ou vont en bénéficier.

S'il est fait appel à un prestataire externe, des garanties contractuelles doivent être prévues : le prestataire a fourni des documents qui attestent de sa conformité au RGPD et d'un niveau de sécurisation adapté à la protection des données à caractère

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<h2>AQ1299</h2> <h3>V1</h3>

personnel (ex: politique de la Sécurité des systèmes d'information (SSI), politique d'assurance sécurité et/ou qualité, politique de protection des données, certification ISO 27001, politique de confidentialité...). Le système choisi par le prestataire ne doit pas conduire à un transfert de données hors Union Européenne sans encadrement juridique.

Dans les cas où des données personnelles peuvent se retrouver sur les postes de travail, alors ces derniers doivent être conformes à la politique du poste de travail INRAE ou de VetAgro Sup.

L'accès aux données se fait via le réseau INRAE ou de VetAgro Sup. En dehors de ces réseaux, le VPN doit être utilisé.

En cas d'incident qui aurait pour conséquence la divulgation des données à caractère personnel concernant ce traitement, le responsable de traitement doit en informer, dès connaissance des faits, le service de la Déléguée à la Protection des Données et les relais de la SSI via le portail Ariane (<https://intranet.inrae.fr/donnees-personnelles/Après-la-mise-en-conformité/Declarer-un-incident>).

La sécurité de ces données comprend également toutes les mesures techniques telles que :

- L'anonymisation des données
- La pseudonymisation des données (remplacer les données directement identifiantes par un n° d'identification) et la gestion sécurisée de la table de correspondance (table faisant le lien entre les données directement identifiantes et le n° d'identification) qui ne doit pas être stockée au même endroit que les données pseudonymisées.
- Le chiffrement des données par un outil de chiffrement préconisé par la SSI, aussi bien sur les serveurs, disques durs des ordinateurs et notamment les portables et les clés USB. À INRAE, des licences du logiciel Zed! sont proposées.
- L'utilisation de mots de passe pour les comptes-utilisateur avec une politique de mot de passe conforme aux recommandations de la CNIL et de l'ANSSI (12 caractères, renouvellements réguliers, etc.).

Il est recommandé de ne pas transférer les données à l'extérieur de la structure. A défaut, ces transferts doivent être sécurisés (chiffrement, utilisation de l'outil File

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<h2>AQ1299</h2> <h2>V1</h2>

Sender de Renater, etc.). Un mot de passe peut être ajouté aux fichiers et communiqué séparément aux partenaires, si le degré de confidentialité/ sensibilité des données l'exige.

Les données à caractère personnel ne peuvent pas être conservées plus longtemps que nécessaire. Une durée de conservation – ou des critères permettant de la définir – doivent donc obligatoirement être spécifiés. Lorsque la finalité du traitement a été atteinte, un archivage pérenne, un effacement ou une anonymisation doivent être organisés. Seule exception : une conservation plus longue pour les traitements ultérieurs en matière de recherche scientifique, si des mesures techniques et d'organisation permettent de garantir les droits des personnes, conformément à l'article 89 du RGPD.

BONNE PRATIQUE : Pour aider à se conformer au RGPD, la DIPS0 a mis les ressources suivantes à disposition :

- Un questionnaire pour déterminer si l'on respecte bien la RGPD est proposé : <https://intranet.inrae.fr/donnees-personnelles/Avant-de-collecter-ou-de-re-utiliser-des-donnees/Bonnes-pratiques-et-situations-frequentes-FAQ/Comment-prendre-en-compte-le-RGPD-dans-les-projets-collaboratifs-europeens-ou-internationaux>
- Des exemples de trame-type de documents pour la mise en conformité avec le RGPD ou le respect de la confidentialité peuvent être trouvés ici : <https://intranet.inrae.fr/donnees-personnelles/Les-demarches-de-mise-en-conformite/Documents-types-disponibles>

6.5. Partage des données

6.5.1. Existe-t-il une obligation de partager ses données (ou à l'inverse une interdiction ou une restriction) ?

Sauf exception, les données produites par une administration dans le cadre de sa mission de service public sont par défaut ouvertes et réutilisables gratuitement. Les exceptions à l'ouverture sont liées à d'éventuelles clauses partenariales limitantes

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<p>AQ1299 V1</p>

et/ou à la nature de certaines données, notamment lorsqu'elles font courir un risque pour autrui.

INRAE organise l'accès aux données qu'il détient selon le principe directeur « **aussi ouvert que possible, aussi fermé que nécessaire** ». Il signifie qu'un accès aux données sans restrictions, libre et gratuit, ne peut s'entendre qu'à la condition préalable d'avoir vérifié le caractère public de ces données et l'absence d'exceptions prévues par la loi (existence de secrets légaux, notamment).

BONNE PRATIQUE : Un logigramme a été créé pour aider à définir si une donnée peut être partagée :

<https://science-ouverte.inrae.fr/fr/les-donnees-et-le-numerique-scientifiques/la-gouvernance-des-donnees/logigramme-aussi-ouvert-que>

En combinant les cadres juridiques en matière de données, un schéma de diagnostic en trois zones est proposé :

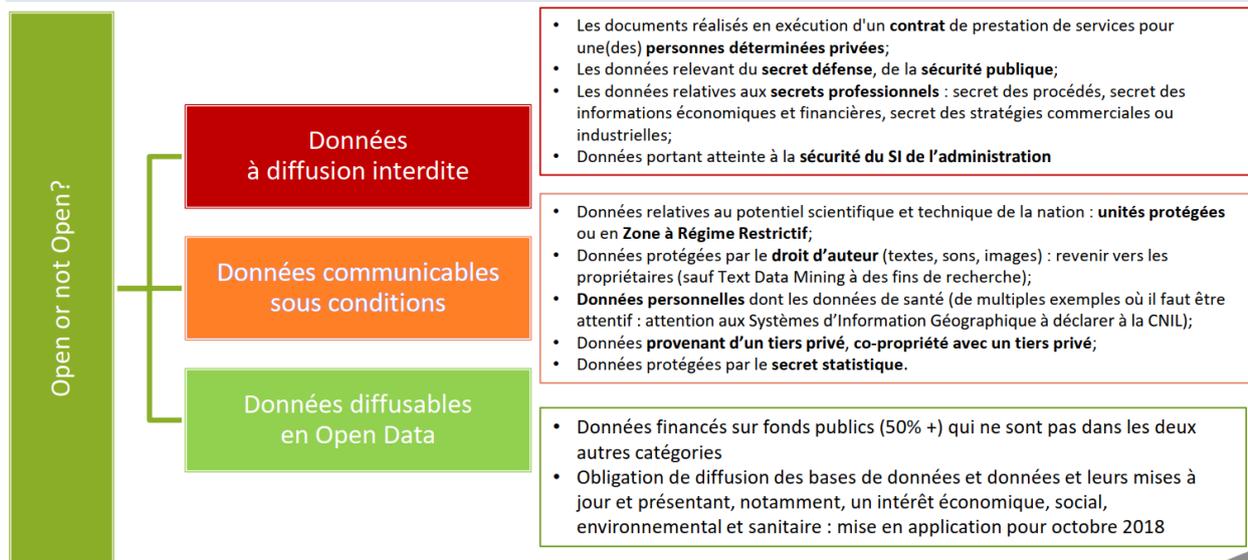


[Note INRAE « Principes pour la gouvernance des données, des codes et des algorithmes » du 2 septembre 2022]

- **Zone verte** : Il n'existe pas de texte juridique (contrat ou loi) qui limite l'usage des données, et les données ne sont pas soumises à une réglementation spécifique. Les conditions de l'accès et de réutilisation gratuites des données par des tiers sont alors réunies.
- **Zone orange** : Il s'agit des cas intermédiaires à approfondir. Par exemple, il existe un texte juridique (contrat ou loi) qui peut limiter l'usage et la gestion des données par des tiers, ou la nature des données peut justifier de limiter l'accès et la réutilisation.
- **Zone rouge** : Il existe un texte juridique (contrat ou loi) qui interdit l'usage et la gestion des données par des tiers, ou elles sont soumises à une réglementation limitante spécifique (données à caractère personnel, RGPD, nature sensible des

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<p style="font-size: 2em; color: #00A68F;">AQ1299</p> <p style="font-size: 2em; color: #00A68F;">V1</p>

données, questions de sécurité nationale, etc.). L'accès et la réutilisation des données sont interdits.



Esther Dzalé / AG CNUE – 05/06/2019

La responsabilité de l'ouverture des données reste portée en premier lieu par le chercheur qui doit en particulier :

- Veiller au cadre éthique (valeurs sociétales, valeurs d'établissement ...) et déontologique (rigueur scientifique, reproductibilité ...) dans lequel s'inscrivent la production et le partage des données.
- Déposer ses données dans des entrepôts disciplinaires ou dans l'entrepôt institutionnel data INRAE
(<https://entrepot.recherche.data.gouv.fr/dataverse/inrae>) afin de leur attribuer un identifiant unique (ex : DOI, URI, etc.) pour permettre de les citer.
- Décrire clairement les données sous forme de métadonnées pour contribuer à les rendre trouvables (Principe FAIR : Findable) et intelligibles pour les ré-utilisateurs potentiels (Principe FAIR : Interoperable). Pour cela, il est conseillé d'y joindre le protocole (où les données confidentielles auront été supprimées)

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<h2>AQ1299</h2> <h3>V1</h3>

ainsi que les modes opératoires utilisés. La qualité des métadonnées et des données doit être garantie par le chercheur.

- Pouvoir valoriser son travail en préservant sa capacité à le poursuivre. Ainsi lorsque les données sont « ouvrables » selon les critères définis par les autres principes, il est possible d'en différer l'ouverture afin de protéger une exploitation scientifique des données par les partenaires du projet ou pour des raisons de sécurité publique. La durée du délai doit toutefois correspondre aux règles éthiques et déontologiques, aux pratiques des communautés, ainsi qu'aux exigences des financeurs, des partenaires et des autorités publiques.

BONNE PRATIQUE : Partager toujours des métadonnées lors du partage des données, dont le protocole et les modes opératoires si nécessaire. Les données et métadonnées seront sans données confidentielles (e.g. nom des personnes hors auteurs du protocole), avec les identifiants d'animaux anonymisés. A noter que les métadonnées sont forcément fournies lors de la rédaction d'un datapaper.

6.5.2. Quelles sont les réutilisations potentielles de ces données ?

Ces données peuvent être réutilisées pour réaliser des méta-analyses, de la modélisation, des analyses faisant appel à de l'agrégation de données ou des analyses temporelles.

6.5.3. Comment les données seront-elles partagées et avec qui ?

Les données en cours d'analyse pourront être partagées avec les agents INRAE et VetAgro Sup et des partenaires identifiés. Elles seront déposées dans des datacenters sécurisés, des bases de données internes ou des sites de partages internes (SharePoint, NextCloud ...).

Une donnée ne sera partagée en accès ouvert que si elle est « achevée » : le traitement de la donnée est finalisé et l'on s'est assuré de la complétude des métadonnées et des documents liés à sa traçabilité. Elle est gage de fiabilité et de qualité de la donnée.

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<h2>AQ1299</h2> <h2>V1</h2>

Suivant le type de données, elle sera alors partagée en open access sur le site recherche.data.gouv.fr et/ou d'autres entrepôts disciplinaires, dans des bases de données publiques ou au travers d'une publication scientifique, afin de leur attribuer un identifiant unique (ex : DOI, URI, etc.) pour permettre leur citation.

6.5.4. Quelle licence utilisée pour protéger ses données ?

Même dans le cas où les données ne sont pas protégées par un droit (droit d'auteur, droit sui generis du producteur de base de données), il est recommandé de les diffuser accompagnées d'un contrat de licence, sous réserve de respecter leur intégrité et de mentionner la source et la date de dernière mise à jour.

Aide au choix des licences : <https://science-ouverte.inrae.fr/fr/les-donnees-et-le-numerique-scientifiques/partager-publier-des-donnees-et-des-codes/comment-choisir>

Le décret 2017-368 (<https://www.legifrance.gouv.fr/eli/decret/2017/4/27/2017-638/jo/texte>) précise les licences gratuites utilisables qui sont l'ODBL et la licence ouverte d'Etalab pour les documents et bases de données.

Exemple de licences existantes :

Licences gratuites	
	<p>ODbL (Open Database License version 1.0) si l'on souhaite pouvoir contrôler les redistributions et les travaux dérivés ; (version française)</p>
	<p>licence Etalab si l'on n'a pas besoin d'un suivi sur le devenir des données et que ces dernières sont essentiellement distribuées en France ;</p>
	<p>Une des licences Creative Commons (CC 4.0). Les licences Creative Commons 4.0 ne prévoient pas l'établissement de redevances mais permettent d'attribuer la production aux auteurs (CC BY) et définir la réutilisation ou la modification des données :</p> <ul style="list-style-type: none"> • de partager dans les mêmes conditions (CC BY-SA) • sans autoriser de modifications : (CC BY-ND)

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<p>AQ1299 V1</p>

	<ul style="list-style-type: none"> • interdire l'utilisation commerciale : (CC BY-NC) <p>de cumuler plusieurs types de restrictions par exemple : (CC BY-NC-SA)</p>
--	--

Il est à noter que la licence ouverte d'Etalab (open access) est compatible avec la licence Creative Commons CC-BY et que la licence ODBL est similaire à la licence Creative Commons CC-BY-SA. Le choix de ces deux licences CC pourra donc s'apprécier en fonction du contexte international et des politiques des entrepôts scientifiques. En revanche les Creative Commons CC-BY-ND (sans modification) et CC-BY-NC (sans usage commercial) sont à éviter dans le contexte de l'open data. Enfin, l'absence d'utilisation de licence (parfois étiquetée CCO) est identique à une ouverture totale des données, c'est-à-dire à les faire entrer dans le domaine public (sans attribution aux auteurs).

BONNE PRATIQUE : Apposer une licence CC-BY sur tous les supports de données produits (preprint, postprint, poster, rapports...)

Pour choisir sa licence des outils existent :

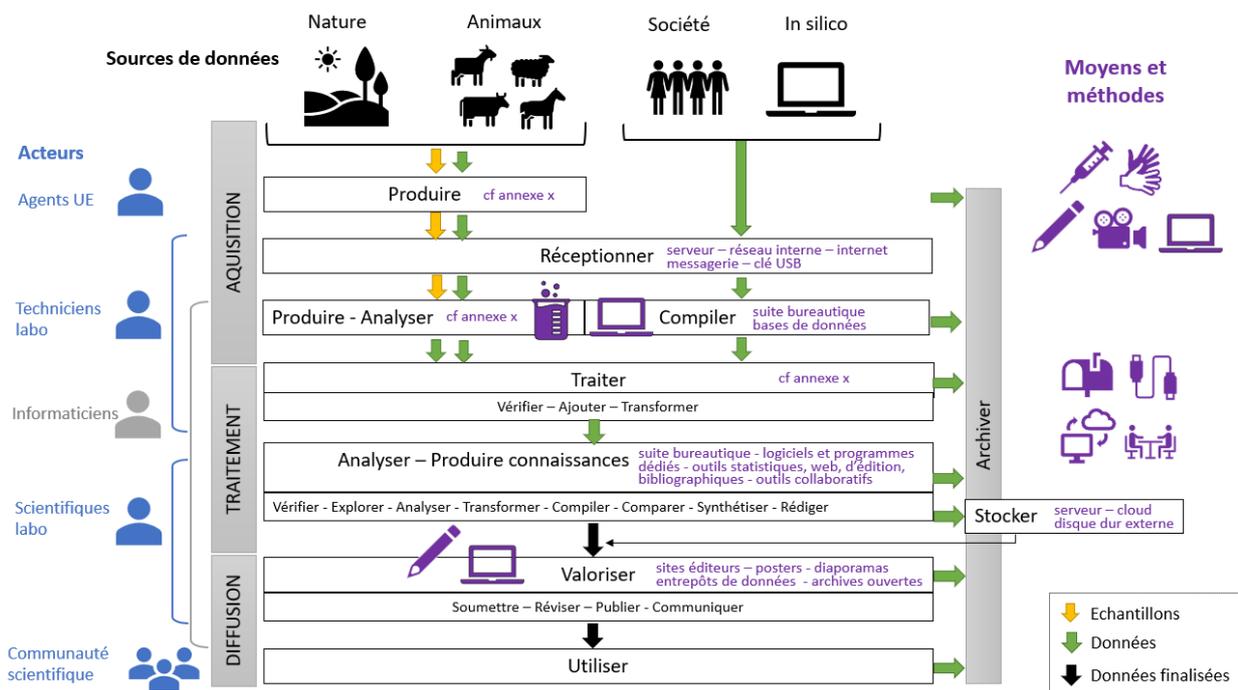
- <http://ufal.github.io/public-license-selector/>
- <http://choosealicense.com/>

Pour les licences logicielles, se référer à l'annexe I.

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<h2>AQ1299</h2> <h2>V1</h2>

6.6. Organisation et documentation des données

6.6.1. Quels méthodes et outils sont utilisés pour acquérir et traiter les données, depuis leur acquisition jusqu'à leur mise à disposition, leur archivage ou leur destruction ?



Groupe Données UMRH INRAE – avril 2022

Passé un « certain temps », les données sont détruites par le gestionnaire des données. Si le support de stockage concernant des données sensibles doit être mis au rebut, il faut impérativement procéder à un formatage à zéro (chaque bit de donnée est remplacé par un zéro).

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

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<p>AQ1299 V1</p>

Les protocoles et modes opératoires sont rédigés selon une trame type et versionnés sur une base de données. Les métadonnées sur le déroulement de l'expérimentation, du dosage et du traitement des données sont écrites dans les cahiers de laboratoire

La rédaction de métadonnées complémentaires est formalisée lors de la préparation de datapaper, à partir des protocoles et des modes opératoires existants.

Parmi les métadonnées, l'usage des ontologies est recommandé pour la description des ontologies. On utilise l'ontologie « Gene Ontology » pour les données transcriptomique et protéomique. Les ontologies ATOL (traits animaux), AHOL (santé/maladie animaux) et EOL (environnement) sont utilisées pour caractériser les données issues des systèmes d'information (SI) SICPA.

En absence d'ontologie, l'utilisation du thesaurus INRAE (<https://consultation.vocabulaires-ouverts.inrae.fr/thesaurus-inrae/fr/>) est suggérée.

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

La documentation supplémentaire nécessaire pour décrire et assurer la « réutilisabilité » (Principe FAIR : Reusable) des données sur le long-terme est fournie à travers les datapapers et les protocoles détaillés (si non inclus dans les métadonnées).

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

Les fichiers de données sont générés en local, puis partagés, gérés et organisés entre les membres d'un projet ou d'une équipe.

Des recommandations de **BONNES PRATIQUES** : sont diffusées pour les aspects :

- Contrôle de version : données gérées sur Git, suivi des versions dans la base Qualité de l'unité (protocoles et modes opératoires)

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<h2>AQ1299</h2> <h3>V1</h3>

- *Convention de nommage -> guide UMRH (AQ1204); convention de nommage dans la base Qualité (protocoles et modes opératoires)*
- *Organisation des fichiers -> structuration commune sur le NAS, structuration par projet sur le NextCloud, guide UMRH (AQ1204)*

6.6.5. Quel est le processus de contrôle qualité des données ?

Au cours de la génération des données, des contrôles qualité sont intégrés dans les modes opératoires (des dosages labo par exemple). Au cours du traitement des données, les contrôles qualité sont complétés par des approches critiques de la qualité des données par expertise et utilisation de méthodes statistiques (normalisation éventuelle, recherche des outliers...)

Une procédure de relecture des modes opératoires est assurée par plusieurs relecteurs avant dépôt sur la base Qualité de l'unité.

6.7. Stockage et sécurité des données

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

Non

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

*Les supports physiques sont des terminaux mobiles de saisie/acquisition (tablette, smartphone ...); ordinateurs (portables et fixes); clés USB; supports papier (dont les cahiers de labo); espaces de stockage sur serveur (NAS); disques durs externes; les clouds institutionnels (Nextcloud INRAE)**

**les clouds de type google-drive sont à exclure.*

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<h2>AQ1299</h2> <h3>V1</h3>

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

La sécurité du transfert des données est assurée par l'utilisation des services institutionnels (par exemple, service de messagerie Outlook INRAE, du serveur interne INRAE (NAS), du cloud INRAE) et par transfert physique via une clé USB. Certains transferts physiques (clé USB, disque dur externe...) font l'objet de chiffrement amont ou aval.

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

Quand l'hébergement est interne à INRAE (ou VetAgro Sup) : c'est la Politique Sécurité des Systèmes d'Information (PSSI) INRAE (ou VetAgro Sup) qui s'applique. Dans les autres cas, demander la politique de sécurité et le plan d'assurance sécurité du partenaire. L'important est de s'assurer de l'adéquation de la PSSI de l'entité hôte avec la réglementation INRAE et avec les impératifs du projet/expérimentation en question.

Voir la charte informatique UMRH (AQ1059v1) pour plus de détails.

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

Les mesures de protection mises en œuvre regroupent : les cahiers de laboratoire, protocoles de recherche, registres de consignation, registres d'anomalies, horodatage des données et résultats, authentification personnelle sur les outils de partage, chiffrement des ordinateurs.

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<h2>AQ1299</h2> <h3>V1</h3>

Voir groupe AQ de l'UMRH. Les correspondants AQ sont les relais dans les équipes autour de ces questions (Mémo du correspondant Qualité AQ1105.)

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

Un plan d'action de sensibilisation est en cours (module d'e-learning INRAE, journée nettoyage des boîtes mail, ...).

6.8. Archivage et conservation des données

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

Toutes les données élaborées et dérivées sont conservées à long terme ainsi que les données brutes utiles (à une ré-exploitation ultérieure)

6.8.2. Sur quelle plateforme d'archivage pérenne seront archivées les données à conserver sur le long terme ?

Sur la plateforme recherche.data.gouv.fr (ou autre dataverse spécifique à un domaine type : proteomeXChange; GEO pour Gene Expression Omnibus); sous la forme de jeux de données déposés et acceptés dans les revues scientifiques (type « supplementary materials ») ou dans des bases de données internes UMRH (e.g. base de données « émissions méthane »).

6.8.3. Sinon, quelles procédures seront mises en place pour la conservation à long terme ?

La conservation à long-terme est assurée par des archivages sur disques durs (pour les données au départ des agents non permanents). Des données en formats libres seront privilégiées.

	<h2>Procédure</h2>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<p>AQ1299 V1</p>

6.8.4. Quelle est la durée de conservation des données ?

Si les données sont concernées par la RGPD alors la durée a été définie lors de la collecte. Sinon, en fonction du PGD du projet. La durée est laissée à l'appréciation du responsable/gestionnaire des données. Une durée d'au moins 10 ans après publication est cependant suggérée.

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

Les coûts associés à la conservation à long terme des données seront supportés par l'Etat français pour le maintien de recherche.data.gouv.fr. Pour les sauvegardes sur disques durs, prévoir des lignes budgétaires dédiées dans les projets (notamment européens).

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<p>AQ1299 V1</p>

Annexe 1 – Licences logicielles

[Extrait de la formation « Le patrimoine immatériel - protection et valorisation » par B. Benifei, 2013.]

6.1. Freeware ou « logiciel gratuit »

Un *freeware* est un logiciel propriétaire distribué gratuitement. Ainsi le code source n'est généralement pas distribué. Il peut également s'accompagner de restrictions d'utilisation (par exemple pas d'usage à but commercial). L'intégralité des droits restent aux propriétaires.

6.2. Licences libres

L'utilisateur d'un logiciel sous licence libre peut :

- Exécuter le programme
- Étudier le code source
- Distribuer des copies
- Améliorer et diffuser librement les modifications

6.2.1. Licences libres sans copyleft

Les licences libres sans *copyleft* sont dites *permissives*. Les œuvres dérivées peuvent être placées sous une licence plus restrictive. Ce sont les licences BSD modifiée (3-clause license), FreeBSD (2-clause licence), MIT, Apache 2.0, CC-BY ou encore CeCILL-B.

6.2.2. Licences Libres avec copyleft

Lorsque la licence est *copyleft*, toute amélioration devra être distribuée sous une licence équivalente (caractère héréditaire/viral/contaminant). Il existe deux types de copyleft : faible/limité et fort.

6.2.3. Copyleft faible ou limité

Les œuvres dérivées doivent être publiées sous une licence équivalente (mais pas les modules détachables). Les bibliothèques sous une licence à *copyleft* faible/limité ne contaminent pas les logiciels les utilisant. Ce sont les licences GNU LGPL, MPL, EPL ou encore CeCILLC.

6.2.4. Copyleft fort

Les œuvres dérivées mais aussi les modules détachables (combinaison d'œuvres) doivent être publiées sous une licence équivalente. Les bibliothèques sous une licence à copyleft fort contaminent les logiciels les utilisant. Ce sont les licences GNU GPL, CC-BY-SA ou encore CeCILL(-A).

6.3. Licences propriétaires

Aussi appelées licences nominatives, elles correspondent à un contrat entre 2 entités.

	<i>Procédure</i>	
Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)	Trame type : AQ485 V5 Créée le 11/02/2020	<i>AQ1299</i> <i>V1</i>

Les licences propriétaires, ne permettent pas l'ensemble des droits du logiciel libre (utiliser, étudier, modifier, dupliquer ou diffuser le logiciel). Le contrat de licence utilisateur final (CLUF) permet de restreindre le droit d'utilisation

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<h2>AQ1299</h2> <h3>V1</h3>

Annexe II – Exemple de réponse de question RGPD (projet européen)

12	What is the legal basis for the processing?	Consent
13	<p>If the legal basis of processing is CONSENT, please identify HOW you receive consent of the individual. (e.g. orally, in written form, electronically etc.)</p> <p>Provide information on the procedure you follow and a copy of the information sheet and consent form.</p> <p>If consent is given electronically, provide a</p>	<p>In written form</p> <p>We include in Annex 1 (at the end of this document) the consent form sent to farmers for signature.</p>
14	<p>Where do you store the personal data? (e.g. on the cloud, on your own servers, on a third party's servers, within EU, outside EU, in an office desk etc.)</p>	on our own servers
15	<p>If personal data is hosted on a third party's servers or the cloud, provide specific information on:</p> <p>the name, type of entity that offers hosting services</p> <p>and whether the third party or cloud provider are GDPR compliant.</p>	Personal data is not hosted in a third party's server

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<p>AQ1299 V1</p>

<p>16</p>	<p><i>For how long do you store/ retain the data?</i></p> <p><i>Under what circumstances do you erase the data?</i></p>	<p><i>The data will be stored until the end of the project development.</i></p> <p><i>Once the carcasses have been classified according to feeding, the personal data</i></p>

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<h2>AQ1299</h2> <h3>V1</h3>

<p>17</p>	<p><i>Do you implement technical and organizational security measures?</i></p> <p><i>Provide a description of the technical and organizational security measures you apply. (e.g. security protocols, encryption, pseudonymization, CCTV, alarm, protection of premises, system security, cyber security, access control, transmission protocols, backups, etc.)</i></p>	<p><i>Yes</i></p> <p><i>The UMRH's NAS space is hosted on INRAE servers hosted in one of the 3 INRAE DataCenters (at Toulouse where the DataCenter is an INRAE building on an INRAE site managed solely by INRAE staff and INRAE service providers, in Ile de France at the TGCC or in Montpellier at the CINES, for these last 2 DataCenters the INRAE infrastructures are hosted in these national DataCenters). In the 3 datacenters, physical access to the servers is subject to access control. Computer access to the servers is secured by INRAE firewalls. NAS spaces access is only possible from INRAE networks (from an INRAE site, or for users connected to the INRAE VPN). Folder permissions are managed by local IT staff on the basis of Active Directory groups. Folders and files in these areas can therefore only be accessed by users authenticated in the INRAE Active Directory and belonging to groups with the appropriate access permissions for folders. In areas order to manage these, server administrators can access the data and the data is not encrypted.</i></p> <p><i>Backups are made automatically using</i></p>
-----------	--	---

	<h1>Procédure</h1>	
<p>Centre Clermont Auvergne-Rhône-Alpes Unité Mixte de Recherche sur les Herbivores (UMRH)</p>	<p>Trame type : AQ485 V5 Créée le 11/02/2020</p>	<p>AQ1299 V1</p>

<p>18</p>	<p><i>Do you keep a 'Record of Processing activities' for the project activities, which contains:</i></p> <ul style="list-style-type: none"> <i>the name and contact details of the controller</i> <i>the purposes of the processing;</i> <i>a description of the categories of data subjects and of the categories of personal data;</i> <i>the categories of recipients to whom the personal data have been or will be disclosed including recipients in third countries or international organizations</i> <i>where applicable, transfers of personal data to a third country or an international organisation,</i> 	