



HAL
open science

PGD du projet "Caractérisation des dangers Clostridium perfringens et Clostridium difficile dans les filières bovine, porcine et volaille en abattoirs"

Olivier Firmesse

► To cite this version:

Olivier Firmesse. PGD du projet "Caractérisation des dangers Clostridium perfringens et Clostridium difficile dans les filières bovine, porcine et volaille en abattoirs". Anses. 2026. <anses-05602265>

HAL Id: anses-05602265

<https://anses.hal.science/anses-05602265v1>

Submitted on 25 Apr 2026

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.



Copyright - All rights reserved

PGD du projet "Caractérisation des dangers Clostridium perfringens et Clostridium difficile dans les filières bovine, porcine et volaille en abattoirs"

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).

Renseignements sur le plan

Titre du plan	PGD du projet "Caractérisation des dangers Clostridium perfringens et Clostridium difficile dans les filières bovine, porcine et volaille en abattoirs"
Livrable	
Version	Version finale
Objet/périmètre du plan	<p>Le projet ClostAbat s'inscrit dans une démarche de production et d'analyse de données microbiologiques, génomiques et environnementales issues de différentes filières d'abattage (bovine, porcine et volaille). Ces données, produites par plusieurs partenaires, nécessitent une gestion structurée et harmonisée tout au long du projet.</p> <p>Le présent Plan de Gestion de Données (PGD) a pour objectif de décrire l'ensemble des structures mises en œuvre pour assurer la collecte, l'organisation, la documentation, le stockage, la sécurisation, le partage et la conservation des données produites. Il vise également à garantir la qualité, la traçabilité et la valorisation des données, dans le respect des exigences réglementaires et des principes de la science ouverte.</p> <p>Ce PGD s'inscrit dans le cadre des recommandations de l'ANR et des principes FAIR (Findable, Accessible, Interoperable, Reusable), afin de faciliter l'accès, la réutilisation et la pérennité des données produites par le projet.</p> <p>Le présent PGD est structuré selon les recommandations de l'ANR et s'organise en six sections principales :</p> <ol style="list-style-type: none">1. Description des données : nature des données produites, méthodes de collecte et formats utilisés2. Documentation et qualité des données : métadonnées, organisation des données et mesures de contrôle qualité3. Stockage et sauvegarde : modalités de stockage, sauvegarde et sécurisation des données4. Exigences légales et éthiques : gestion des données personnelles, propriété intellectuelle et aspects réglementaires5. Partage des données et conservation : modalités de diffusion, accès, archivage et conservation à long terme6. Responsabilités et ressources : organisation de la gestion des données et moyens mobilisés

Cette structuration permet d'assurer une gestion cohérente et transparente des données tout au long du cycle de vie du projet.

Domaines de recherche (selon classification de l'OCDE)

Health sciences, Biological sciences (Natural sciences), Biological sciences

Langue

fra

Date de création

2022-03-25

Date de dernière modification

2026-04-13

Renseignements sur le projet**Titre du projet**

Caractérisation des dangers Clostridium perfringens et Clostridium difficile dans les filières bovine, porcine et volaille en abattoirs

Acronyme

ClostAbat

Résumé

Clostridium perfringens et Clostridioides difficile sont des bactéries pathogènes majeures impliquées dans les infections humaines. C. perfringens est de plus l'un des dangers bactériens identifiés dans les Guides de Bonnes Pratiques d'Hygiène et d'application des principes HACCP du maillon abattage-découpe des filières bovine, porcine et volaille. Les animaux étant porteurs sains de ces deux espèces dans leur tube digestif, les viandes animales sont des cibles privilégiées pour expliquer la survenue d'épisodes toxiques impliquant ces pathogènes. Sans que l'adéquation et la pertinence des mesures de maîtrises mises en place en abattoir et en découpe ne soient remises en cause, peu ou pas de données sont aujourd'hui disponibles en France sur l'importance de ces deux dangers au maillon abattoir et atelier de découpe. Le projet ClostAbat réunit des partenaires académiques et des instituts techniques avec un large éventail de compétences en sciences des aliments et en clinique humaine afin de développer des connaissances et des outils d'aide à la décision pour permettre aux filières d'avoir une information objective sur l'importance de ces dangers en abattoirs/ateliers de découpe. Ce projet permettra : i) de déterminer le statut de contamination d'échantillons de lots de viandes provenant des filières porcines, bovines et volailles à l'égard des dangers C. perfringens et C. difficile à l'abattoir et d'envisager la corrélation entre les prévalences de portage digestif, de l'environnement de production et de contamination des carcasses et produits de découpe. ii) de déterminer le potentiel pathogène des souches isolées pour l'homme ; iii) d'identifier des marqueurs épidémiologiques permettant la traçabilité des isolats en fonction des filières animales ; iv) de définir leur persistance en abattoir et la corrélation potentielle entre leur présence et l'abondance d'autres taxons bactériens dans les populations de surface ; v) de déterminer l'importance relative des différentes voies de transmission (animale ou environnementale) et leur implication possible dans les maladies infectieuses chez l'homme ; vi) de transférer les méthodologies utilisées pour le traitement des données aux instituts techniques pour la caractérisation et la surveillance des deux pathogènes dans les différents secteurs de la viande ; vii) d'établir des recommandations en rédigeant des lignes directrices décrivant les

stratégies à mener par les opérateurs afin de réduire efficacement leur implication dans la contamination et leur impact sur la Santé Publique ; viii) de proposer une stratégie exploratoire alternative pour la gestion de ces dangers par le traitement à la lumière LED et évaluer son impact sur la qualité de la viande. Le projet comporte 5 programmes de travail (WP) : Le WP1 concerne l'organisation de la collecte des échantillons dans les abattoirs en fonction des spécificités de chaque secteur, le WP2 vise à déterminer les fréquences de contamination dans les trois filières viande au regard des dangers C. perfringens et C. difficile, le WP3 a pour objectif de déterminer le potentiel pathogène des souches isolées pour l'Homme, le WP4 vise à identifier l'importance relative des filières au regard des deux dangers et le WP5 vise à investiguer la possibilité d'utiliser la lumière LED combinée à l'humidité relative/température pour détruire les spores de C. perfringens et C. difficile. Ainsi, les opérateurs des filières viandes seront mieux à même de justifier l'adéquation et la pertinence des mesures de maîtrise mises en œuvre à l'égard des dangers liés à C. perfringens et C. difficile, et, si nécessaire, de les réexaminer et de diffuser des pratiques de prévention retenues.

Sources de financement

Financier	Nom du financier	Agence Nationale de la Recherche
	Identifiant	https://ror.org/00rbzpz17
	Type d'identifiant	ROR
Identifiant du financement		ANR-21-CE21-0007
Statut du financement		Approuvé

Date de début

2022-01-01

Date de fin

2025-12-31

Partenaires

Nom	IP-CNR Bactéries anaérobies et Botulisme
Nom	DMU APHP.Sorbonne : Biologie et Génomique Médicales
Identifiant RNSR	201923346Y
Nom	PROCEDES ALIMENTAIRES et MICROBIOLOGIQUES - UMR_MA 2102
Identifiant RNSR	200715421L
Nom	Institut de lélevage

Nom	IFIP-Institut du Porc
Nom	ANSES - Laboratoire de Ploufragan-Plouzané-Niort
Identifiant RNSR	195723283Z
Nom	PROCEDES ALIMENTAIRES et MICROBIOLOGIQUES - UMR_MA 2102
Identifiant RNSR	200715421L
Nom	IFIP-Institut du Porc
Nom	Institut de l'élevage
Nom	ANSES - Laboratoire de sécurité des aliments, sites de Maisons-Alfort et de Boulogne-sur-Mer
Identifiant RNSR	201023301N
Nom	DMU APHP.Sorbonne : Biologie et Génomique Médicales
Identifiant RNSR	201923346Y
Nom	ANSES - Laboratoire de Ploufragan-Plouzané-Niort
Identifiant RNSR	195723283Z
Nom	IP-CNR Bactéries anaérobies et Botulisme

Produits de recherche :

1. WP1-Management du projet (Texte)
2. WP2-Campagnes de prélèvements (Collection)
3. WP3-Détection de *C. perfringens* et *C. difficile* dans les échantillons (Collection)
4. WP4-Diversité génétique et toxinotype des souches de *C. perfringens* et *C. difficile* (Jeu de données)
5. WP5-Structure des populations et attribution de source (Jeu de données)
6. WP6-Evaluation et proposition de mesures de maîtrise (Jeu de données)

Contributeurs

Nom	Affiliation	Rôles
FIRMESE Olivier - https://orcid.org/0000-0003-2415-2802	ANSES - Laboratoire de sécurité des aliments, sites de Maisons-Alfort et de Boulogne-sur-Mer - 201023301N	<ul style="list-style-type: none"> • Coordinateur du projet • Personne contact pour les données (WP6, WP4, WP5, WP1, WP2, WP3) • Responsable du plan de gestion de données

PGD du projet "Caractérisation des dangers Clostridium perfringens et Clostridium difficile dans les filières bovine, porcine et volaille en abattoirs"

WP1-Management du projet

1. Description des données et collecte ou réutilisation de données existantes

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Les données du WP1 (management du projet) sont générées dans le cadre des interactions entre les partenaires, en amont et à l'issue des différentes réunions organisées tout au long du projet.

Une réunion de lancement est organisée au démarrage du projet, suivie de réunions semestrielles. En amont de ces réunions, chaque partenaire transmet un rapport sous forme de présentation, comprenant les résultats obtenus, le plan de travail pour les 12 mois suivants ainsi que les actions de diffusion réalisées. Lors des réunions, ces éléments sont présentés et discutés collectivement. À l'issue de chaque réunion, un compte rendu synthétisant les décisions prises est rédigé par le coordinateur puis transmis à l'ensemble des participants pour validation.

Des réunions complémentaires, organisées majoritairement en visioconférence afin de limiter les déplacements et leur impact environnemental, sont mises en place en fonction des besoins, notamment au démarrage des sous-tâches ou pour traiter des questions spécifiques. Ces réunions sont pilotées par les responsables de tâches ou les partenaires concernés, avec participation du coordinateur. Un compte rendu est rédigé par l'organisateur et soumis à validation des participants.

Afin de faciliter les échanges, le partage des données et la coordination du projet, un espace de travail sécurisé dédié est mis en place sur la plateforme RESANA. Cet espace, accessible uniquement aux partenaires, permet le dépôt, la consultation et l'échange des documents produits (rapports, présentations, comptes rendus, données de suivi).

Aucune réutilisation de données préexistantes n'est réalisée dans le cadre du WP1.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Les données produites dans le WP1 sont principalement des données de gestion de projet, comprenant :

- des fichiers tabulaires (bases de données, tableurs)
- des documents textuels (rapports, comptes rendus, présentations, notes de travail)
- des supports visuels (présentations, images)

Les formats utilisés sont standards et ouverts afin de garantir la lisibilité et la réutilisation à long terme (XLSX, PDF, DOCX, JPEG).

Les volumes de données générés sont limités et ne présentent pas de contrainte particulière en termes de stockage.

2. Documentation et qualité des données

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

Les données sont organisées selon une nomenclature standardisée permettant leur identification et leur traçabilité. Chaque fichier comporte a minima :

- la date

- le type de document ou de réunion
- l'identifiant du projet
- les initiales du rédacteur (le cas échéant)

Exemple : 2023_01_15_ClostAbat_WP1_Presentation_OF

Cette organisation garantit une identification rapide et un classement cohérent des données.

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Les documents produits (rapports, comptes rendus) font l'objet d'une validation collective par les partenaires du projet. Un délai de 15 jours est accordé aux partenaires pour proposer des modifications. En l'absence de retour dans ce délai, les documents sont considérés comme validés.

Toute modification ultérieure nécessite une validation collective.

3. Stockage et sauvegarde pendant le processus de recherche

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Les données et métadonnées du WP1 sont stockées :

- sur l'espace collaboratif sécurisé RESANA
- sur les serveurs du laboratoire coordinateur

Des sauvegardes régulières (quotidiennes sur les serveurs) sont mises en place afin de garantir l'intégrité et la pérennité des données.

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

L'accès aux données est sécurisé :

- authentification par identifiant et mot de passe
- accès restreint aux seuls partenaires du projet
- gestion des droits (lecture, modification, suppression)

La duplication des supports de stockage permet de prévenir les pertes de données en cas d'incident.

Les données sensibles sont accessibles uniquement aux personnes autorisées.

4. Exigences légales et éthiques, codes de conduite

4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Les données à caractère personnel, lorsqu'elles existent, sont anonymisées conformément à la réglementation en vigueur (RGPD).

4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Les droits de propriété intellectuelle et les modalités de gestion des données sont définis dans l'accord de consortium signé par l'ensemble des partenaires.

Conformément à la loi n°2016-1321 pour une République numérique, les données issues d'une recherche financée sur fonds publics ont vocation à être rendues accessibles, sous réserve de l'absence de contraintes juridiques ou de confidentialité.

4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Le projet ne soulève pas de problématiques éthiques spécifiques.

5. Partage des données et conservation à long terme

5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?

L'ensemble des données sont partagées entre les partenaires du projet via l'espace sécurisé RESANA.

Les prélèvements réalisés lors des campagnes de prélèvement sont détruit après analyse.

Les données concernant le nom des établissements participants resteront confidentielles et une anonymisation des données de séquences humaines sera respectée.

Les données de valorisation (avancements du projet, affiches, conférences, articles ...) seront accessibles progressivement, au fur et à mesure de leur intégration sur le site internet tout au long du projet. Nous souhaitons rendre ces données accessibles le plus rapidement possible afin de rendre visible l'état d'avancement du projet et dans les principes de la science ouverte.

L'ensemble des données collectées non confidentielles dans le cadre de ce projet seront disponibles en libre accès après l'embargo de publication finale des données traitées comme indiqué dans l'accord de consortium.

5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

Pour le moment, les partenaires du projet s'engagent à conserver l'ensemble des données/métadonnées produites sur les serveurs de leurs instituts respectifs pendant toute la durée du projet et à minima pendant 3 années supplémentaires.

Tous les cahiers de laboratoire papier seront conservés et archivés.

Les données génomiques déposées dans Genbank seront préservées dans le cadre des missions du NCBI (<https://www.ncbi.nlm.nih.gov/genbank/>).

Plus particulièrement, pour les partenaires :

Anses Maisons-Alfort :

L'ensemble des souches de *C. perfringens* envoyées par les partenaires seront conservées à -80°C.

Les données collectées sont conservées sur le réseau de l'Anses.

Anses Ploufragan (filière volailles) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés en DASRI lorsque la mise en collection des isolats est terminée. Les isolats sont stockés au surgélateur. Pour *C. perfringens*, deux tubes sont préparés par isolat : un est stocké à Ploufragan, le second est envoyé à Maisons-Alfort. Pour *C. difficile*, les isolats sont repiqués sur ChromID à partir du tube

stocké au surgélateur puis envoyé à l'hôpital Saint-Antoine (WP3).

Les prélèvements destinés à l'études des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant extraction et séquençage (WP6).

les données de séquençage seront déposées dans des bases publiques (SRA).

Les données collectées sont conservées sur le réseau de l'Anses.

Institut Pasteur :

Les données de séquençage des souches de *C. perfringens* et *C. difficile* sont stockées sur le réseau de l'Institut Pasteur.

Hôpital Saint-Antoine :

Archivage papier (feuilles de paillasse + CR) dans le bureau du CNR *Clostridioides difficile* au sein de l'Hôpital St Antoine.

Archivage informatique (réseau APHP + disque dur externe pour les fichiers .fsa + Résana + site Pasteur)

UMR PAM :

L'ensemble des données produites seront conservées sur le long terme par un cahier de laboratoire et un disque dur externe. Le cahier de laboratoire et le disque dur externe seront conservés à vie dans des armoires fermant à clé dans l'unité de recherche.

IDELE (filiale bovine) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés lorsque la confirmation des souches isolées (*C. perfringens* par ANSES et *C. difficile* par Hôpital Saint-Antoine) a été réalisée. Les souches isolées sont conservées à -80°C en parallèle de l'envoi à l'ANSES Maisons-Alfort et à l'Hôpital Saint-Antoine (WP3). Les prélèvements destinés à l'études des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant envoi à l'ANSES de Ploufragan (WP6).

Pour le WP3, les souches potentielles issues de la filière bovine sont identifiées de la manière suivante (commune avec les autres partenaires) : CBAT-P-B-XXX (ClostAbat - Perfringens - Bovin - numéro de 1 à XXX) et CBAT-D-B-XXX (ClostAbat - Difficile - Bovin - numéro de 1 à XXX).

IFIP (filiale porcine) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés après traitement lors de chaque campagne. Les souches isolées sont conservées à -80°C en parallèle de l'envoi à l'ANSES et à l'hôpital Saint-Antoine (WP3).

Les prélèvements destinés à l'études des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant envoi à l'ANSES Ploufragan (WP6).

Les données obtenues sont conservées dans l'espace de stockage des données de l'IFIP par archivage

5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

Les outils logiciels pour accéder et utiliser les données/métadonnées du projet sont standards et ouverts pour faciliter le partage et la réutilisation à long terme des données.

Néanmoins, certains logiciels utilisés pour les analyses sont restreints d'accès :

- GeneMapper : logiciel installé sur les ordinateurs du CNR *Clostridioides difficile*. Code de connexion (login + mot de passe) unique pour l'équipe du CNR

- Webribo pour la lecture des fichiers .fsa (données bruts) : site internet. Chaque nouvel utilisateur doit se créer son compte. Pas de visualisation possible des données des autres comptes (mais possibilité de retélécharger le fichier .fsa de la souche issu du séquenceur).

- BioNumerics : logiciel d'analyse des profils de typage moléculaire M13 pour les souches de *C. perfringens*.

5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?

Il n'est pas envisagé actuellement d'attribuer systématiquement de DOI à la production d'un jeu de données. Cette attribution reste à la

charge du responsable scientifique de chaque partenaire du projet réalisant des analyses.

Chaque donnée est identifiée par une numérotation particulière défini pour chaque partenaire en charge des prélèvements.

A savoir :

Anses Maisons-Alfort :

Dans le cadre du WP1, afin qu'ils soient faciles à retrouver et identifier, le nom du fichier de chaque rapport comporte la date et le type de réunion et dans le cadre d'une présentation les initiales de la personne qui l'a réalisé (exemple : 2023_01_15_PresentationWP1_OF).

Les données sont référencées par un code défini en interne par chaque partenaire qui permet de les identifier et les retrouver facilement. A titre d'exemple dans le WP1, le nom du fichier de chaque rapport comportera la date et le type de réunion et dans le cadre d'une présentation les initiales de la personne qui l'a réalisé (exemple : 2023_01_15_PresentationWP1_OF).

Les publications produites seront identifiées par un DOI.

Anses Ploufragan (filière volailles) :

Pour le WP2, un N° de série a été attribué au projet: 23TP01-AA (où AA est le N° de l'analyse, 23TP01-01 signifie que c'est la première mise en analyse du projet Clost'Abat). Les échantillons sont numérotés avec le code 23TP01-XX où XX est le N° de l'échantillon. Cette numérotation est celle décrite dans les documents qualité de l'unité HQPAP (PPN/PS/0200).

Hôpital Saint-Antoine :

Dans le cadre du WP4, un numéro interne de suivi de souches est attribué à chaque souche avérée *Clostridioides difficile*.
Exemple : CD23-037, avec :

- CD : pour *Clostridioides difficile*
- 23 : les deux derniers chiffres de l'année en cours
- 037 : numéro d'incrémentation des souches du CNR.

Ce numéro suit la souche : tableau Excel, site Pasteur, feuilles de paillasse, fichier .fsa, compte rendu ...

UMR PAM :

Chaque nouvelle expérience est identifiée par un code selon la condition testée. Par exemple, pour les essais d'humidité relative sur les cellules végétatives, le code suivant a été utilisé : LS_CP_HR_CV001

- LS : nom de l'expérimentateur
- CP : nom de l'espèce bactérienne
- HR : humidité relative
- CV : cellules végétatives
- 001 : numéro d'expérience référencé à une date dans le cahier de laboratoire et dans un tableur

Les publications produites seront identifiées par un DOI.

IFIP (filière porcine) :

Pour le WP2, les prélèvements sont identifiés selon une nomenclature standardisée. Le n° d'échantillon est composé du code projet CBAT suivi du code Abattoir Ab, HP ou TL, du n° de campagne 1, 2 ou 3, du numéro de lot prélevé le jour de la campagne, de 1, à 4, du numéro de porc prélevé, de 1, à 5 et du type de prélèvement, F pour fèces, C pour carcasses et P pour le muscle Poitrine : CBAT-HP1-1-4-P.

pour les prélèvements d'environnements, les codes sont comme suit : Nom du projet CBAT, code abattoir, n° de campagne, A pour abattoir et D pour découpe, air 1 à air 3 pour les échantillons d'air, E1 à E5 pour les échantillons de surface: CBAT-HP1-D-E1 ou CBAT-HP1-D-air1.

Pour les souches détectées et stockées, les numéros suivent la même nomenclature que le prélèvement auquel est ajouté col1 à col5 pour colonie 1 à 5 : CBAT-Ab1-3-3-C-col4

IDELE (filière bovine) :

Pour le WP2, les prélèvements de chaque campagne sont identifiés de la même manière, avec un code spécifique pour chaque campagne. Chaque campagne est numéroté de 1 à 8 avec la localisation du prélèvement. Les fèces sont identifiés avec F, les carcasses avec C, les muscles avec M, les environnements abattoir EA, les environnement découpe ED.

Par exemple, l'échantillon de fèces de l'animal 1 de la campagne numéro 1 est noté : C1F1 : C1 (campagne 1, fèces animal numéro 1). Et ainsi de suite ...

6. Responsabilités et ressources en matière de gestion des données

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

La gestion des données (saisie, production, stockage, sauvegarde) est assurée par les responsables scientifiques de chaque partenaire, conformément aux règles de leurs institutions.

Le coordinateur du projet est responsable de la mise en œuvre et du suivi du PGD.

6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

Le financement du stockage local des données est assuré par l'institut hôte, l'Anses qui centralisera l'ensemble des métadonnées générées dans le cadre du projet.

L'accessibilité et la réutilisation des données ouvertes seront assurées par prise de contact avec le coordinateur et les responsables scientifiques de chaque WP, leurs adresses électroniques se trouvant sur le site internet dédié au projet. Du temps dédié à la gestion des données est prévu, notamment par le coordinateur, afin de garantir :

- la qualité des données
- leur documentation
- leur conformité aux principes FAIR (Findable, Accessible, Interoperable, Reusable)

L'utilisation de formats ouverts et de standards partagés contribue à assurer l'interopérabilité et la réutilisation des données.

WP2-Campagnes de prélèvements

1. Description des données et collecte ou réutilisation de données existantes

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Les données du WP2 sont issues de campagnes de prélèvements réalisées dans des abattoirs des filières bovine, porcine et volaille.

Les campagnes sont conduites selon des protocoles harmonisés entre partenaires afin d'assurer la comparabilité des données. Chaque campagne comprend un nombre défini de prélèvements effectués sur différents types de matrices : fèces, carcasses, produits carnés, surfaces environnementales et air.

Les prélèvements sont réalisés directement en abattoir par les partenaires selon des procédures standardisées. Les échantillons sont ensuite conditionnés, transportés et analysés dans les laboratoires partenaires.

Aucune réutilisation de données préexistantes n'est prévue dans ce WP.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Les données produites incluent :

- données de prélèvement (localisation, type d'échantillon, filière)
- données de traçabilité (campagne, lot, animal)
- données environnementales (zone de prélèvement, type de surface, air)

Formats :

- fichiers tableurs (XLSX, CSV)
- fiches de prélèvement (PDF)

Les volumes de données sont modérés mais répétés à chaque campagne.

2. Documentation et qualité des données

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

Un protocole d'échantillonnage commun (quand, où et comment prélever ?) à tous les partenaires impliqués dans les campagnes de prélèvements a été défini afin d'harmoniser au mieux les méthodes pour limiter au maximum un éventuel effet "prélèvement" dans les résultats et pouvoir ainsi comparer les filières.

Les prélèvements réalisés lors de chaque campagne seront répertoriés par filière dans un fichier numérique type tableur. Ce fichier sera mis en commun entre les partenaires sur l'espace de travail RESANA consacré au projet ClostAbat.

Afin de décrire au mieux les données générés lors des prélèvements, les informations suivantes sur les animaux prélevés ont été recueillies :

IFIP (filière porcine) :

En plus des prélèvements effectués, les informations suivantes sur les animaux prélevés ont été recueillies :

- Numéro de tuerie
- Numéro de TVA
- Département d'élevage

Les procédures de nettoyage/désinfection et les fiches de sécurité des produits utilisés ont été fournis par les professionnels des abattoirs.

IDELE (filière bovine) :

En plus des prélèvements effectués, les informations suivantes sur les animaux prélevés ont été recueillies :

- Numéro de tuerie
- Catégorie de l'animal (vache, génisse, taureau, jeune bovin, bœuf)
- Race de l'animal
- Date de naissance
- Département de naissance et département d'élevage
- Type d'abattage (classique/rituel)

Les procédures de nettoyage/désinfection et les fiches de sécurité des produits utilisés ont été fournis par les professionnels des abattoirs.

Anses Ploufragan (filière volailles) :

La fiche ICA (information sur la chaîne alimentaire) a été récupérée pour chaque lot. Cette fiche donne les informations suivantes:

- Nom de l'élevage et N° EDE
- organisation de production
- N° bâtiment
- Couvoir d'origine
- Type de production
- Nombre d'animaux mis en place et date de mise en place
- Nombre d'animaux lors de l'enlèvement
- Chargement (kg/m²)
- Programme alimentaire
- Poids vif moyen
- Mortalité
- Présence d'un traitement antibiotique
- Pathologies, accidents

Les procédures de nettoyage/désinfection et les fiches de sécurité des produits utilisés ont été fournis par les professionnels des abattoirs.

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Protocoles standardisés entre partenaires

Formation des opérateurs

Contrôle positif en parallèle du traitement des échantillons (T+ pour chaque pathogène)

Fiche de suivi et traçabilité des échantillons (n°lot, réactif, fournisseur, suivi de température, matériels utilisés ...)

Vérification et étalonnage des appareils de mesure, EIL (RAEMA), procédures qualité internes

3. Stockage et sauvegarde pendant le processus de recherche

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Les données et métadonnées seront stockées et sauvegardées tout au long du projet sur des cahiers de terrain lors des campagnes de prélèvements, des cahiers de laboratoire repertoriés et stockés dans des armoires fermant à clé, sur les ordinateurs des membres du projet, sur des disques durs externes, sur les serveurs des instituts partenaires du projet et sur un serveur intranet de l'Anses uniquement dédié aux données générées lors du projet et accessible uniquement par les membres du projet (en lecture seule sauf pour les responsables scientifiques de chaque partenaire).

IFIP (filière porcine) :

Enregistrement des échantillons/heure de mesure des sondes déposées : 1 fichier type tableur de traçabilité par campagne pour numéroter les échantillons

Traitement des échantillons : sur fichier type tableur (différent de l'enregistrement)

Mesures pendant les campagnes : sur papier puis stockage sur Cloud (enregistrement des échantillons)

Stockage :

OneDrive dans un datacenter de Microsoft en France

IDELE (filière bovine) :

Les informations de traçabilité (informations sur les prélèvements, heure de dépôt des sondes, etc.) sont collectées sur papier dans l'abattoir puis sont ensuite enregistrées sur un fichier type tableur dédié. Le traitement des échantillons et les résultats par campagne sont traités sur un fichier type tableur séparé. Enfin, tous les résultats sont saisis et enregistrés sur un fichier partagé RESANA avec une copie sur un fichier type tableur en interne.

Stockage :

Les données sont stockées sur le disque dur de l'ordinateur ainsi que sur le réseau interne à l'IDELE (sauvegarde automatique plusieurs fois par jour).

Anses Ploufragan (filière volailles) :

Les informations de traçabilité (informations sur les prélèvements, heure de dépôt des sondes, etc.) sont collectées sur papier dans l'abattoir puis sont ensuite enregistrées dans un fichier type tableur dédié. Le traitement des échantillons et les résultats par campagne sont tracés dans les fiches de paillasse stockées dans un classeur et les résultats sont saisis dans un fichier type tableur, enregistré sur le réseau de l'Anses dans un dossier dont l'accès est réservé aux personnes impliquées dans le projet au sein du laboratoire de Ploufragan-Plouzané-Niort. Enfin, tous les résultats sont saisis et enregistrés sur un fichier partagé RESANA .

Stockage :

Les données sont stockées sur le réseau de l'Anses (copie du réseau quotidienne sur un autre serveur).

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

La diversité des supports de sauvegarde des données mise en place permet d'assurer la sécurité des données. Rien ne s'opposera à leur récupération en cas d'incident.

Les données collectées présentes sur les ordinateurs des partenaires du projet sont protégées par un mot de passe.

Les données stockées sur le serveur intranet de l'Anses sont accessibles (en lecture seule) uniquement à tous les membres du projet au cours du processus de recherche. Elles ne sont modifiables que par les responsables scientifiques de chaque partenaire.

En ce qui concerne la protection des données, les données ne sont pas considérées comme sensibles. Pas de protection particulière.

IFIP (filière porcine) :

Stockage sécurisé par double authentification pour l'accès. Les membres du projet peuvent avoir accès par invitation et inscription dans notre Azur Directory.

IDELE (filière bovine) :

Les données sont protégées par le mot de passe de l'ordinateur de chaque utilisateur. Elles ne sont disponibles que pour les personnes impliquées dans le projet ClostAbat.

Anses Ploufragan (filière volailles) :

Les données sont stockées sur le réseau de l'Anses avec un accès restreint aux personnes impliquées dans le projet au sein du laboratoire de l'Anses de Ploufragan-Plouzané-Niort.

4. Exigences légales et éthiques, codes de conduite

4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Les données à caractère personnel, lorsqu'elles existent, sont anonymisées conformément à la réglementation en vigueur (RGPD).

4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Les droits de propriété intellectuelle et les modalités de gestion des données sont définis dans l'accord de consortium signé par l'ensemble des partenaires.

Conformément à la loi n°2016-1321 pour une République numérique, les données issues d'une recherche financée sur fonds publics ont vocation à être rendues accessibles, sous réserve de l'absence de contraintes juridiques ou de confidentialité.

4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Le projet ne soulève pas de problématiques éthiques spécifiques.

5. Partage des données et conservation à long terme

5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?

L'ensemble des données sont partagées entre les partenaires du projet via l'espace sécurisé RESANA.

Les prélèvements réalisés lors des campagnes de prélèvement sont détruit après analyse.

Les données concernant le nom des établissements participants resteront confidentielles et une anonymisation des données de séquences humaines sera respectée.

Les données de valorisation (avancements du projet, affiches, conférences, articles ...) seront accessibles progressivement, au fur et à mesure de leur intégration sur le site internet tout au long du projet. Nous souhaitons rendre ces données accessibles le plus rapidement possible afin de rendre visible l'état d'avancement du projet et dans les principes de la science ouverte.

L'ensemble des données collectées non confidentielles dans le cadre de ce projet seront disponibles en libre accès après l'embargo de publication finale des données traitées comme indiqué dans l'accord de consortium.

5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

Pour le moment, les partenaires du projet s'engagent à conserver l'ensemble des données/métadonnées produites sur les serveurs de leurs instituts respectifs pendant toute la durée du projet et à minima pendant 3 années supplémentaires.

Tous les cahiers de laboratoire papier seront conservés et archivés.

Les données génomiques déposées dans Genbank seront préservées dans le cadre des missions du NCBI (<https://www.ncbi.nlm.nih.gov/genbank/>).

Plus particulièrement, pour les partenaires :

Anses Maisons-Alfort :

L'ensemble des souches de *C. perfringens* envoyées par les partenaires seront conservées à -80°C.

Les données collectées sont conservées sur le réseau de l'Anses.

Anses Ploufragan (filière volailles) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés en DASRI lorsque la mise en collection des isolats est terminée. Les isolats sont stockés au surgélateur. Pour *C. perfringens*, deux tubes sont préparés par isolat : un est stocké à Ploufragan, le second est envoyé à Maisons-Alfort. Pour *C. difficile*, les isolats sont repiqués sur ChromID à partir du tube stocké au surgélateur puis envoyé à l'hôpital Saint-Antoine (WP3).

Les prélèvements destinés à l'études des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant extraction et séquençage (WP6).

les données de séquençage seront déposées dans des bases publiques (SRA).

Les données collectées sont conservées sur le réseau de l'Anses.

Institut Pasteur :

Les données de séquençage des souches de *C. perfringens* et *C. difficile* sont stockées sur le réseau de l'Institut Pasteur.

Hôpital Saint-Antoine :

Archivage papier (feuilles de paillasse + CR) dans le bureau du CNR *Clostridioides difficile* au sein de l'Hôpital St Antoine.

Archivage informatique (réseau APHP + disque dur externe pour les fichiers .fsa + Résana + site Pasteur)

UMR PAM :

L'ensemble des données produites seront conservées sur le long terme par un cahier de laboratoire et un disque dur externe.

Le cahier de laboratoire et le disque dur externe seront conservés à vie dans des armoires fermant à clé dans l'unité de recherche.

IDELE (filière bovine) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés lorsque la confirmation des souches isolées (*C. perfringens* par ANSES et *C. difficile* par Hôpital Saint-Antoine) a été réalisée. Les souches isolées sont conservées à -80°C en parallèle de l'envoi à l'ANSES Maisons-Alfort et à l'Hôpital Saint-Antoine (WP3). Les prélèvements destinés à l'études des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant envoi à l'ANSES de Ploufragan (WP6).

Pour le WP3, les souches potentielles issues de la filière bovine sont identifiées de la manière suivante (commune avec les autres partenaires) : CBAT-P-B-XXX (ClostAbat - Perfringens - Bovin - numéro de 1 à XXX) et CBAT-D-B-XXX (ClostAbat - Difficile - Bovin - numéro de 1 à XXX).

IFIP (filière porcine) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés après traitement lors de chaque campagne. Les souches isolées sont conservées à -80°C en parallèle de l'envoi à l'ANSES et à l'hôpital Saint-Antoine (WP3).

Les prélèvements destinés à l'études des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant envoi à l'ANSES Ploufragan (WP6).

Les données obtenues sont conservées dans l'espace de stockage des données de l'IFIP par archivage

5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

Les outils logiciels pour accéder et utiliser les données/métadonnées du projet sont standards et ouverts pour faciliter le partage et la réutilisation à long terme des données.

Néanmoins, certains logiciels utilisés pour les analyses sont restreints d'accès :

- GeneMapper : logiciel installé sur les ordinateurs du CNR *Clostridioides difficile*. Code de connexion (login + mot de passe) unique pour l'équipe du CNR

- Webribo pour la lecture des fichiers .fsa (données brutes) : site internet. Chaque nouvel utilisateur doit se créer son compte. Pas de visualisation possible des données des autres comptes (mais possibilité de retélécharger le fichier .fsa de la souche issu du séquenceur).

- BioNumerics : logiciel d'analyse des profils de typage moléculaire M13 pour les souches de *C. perfringens*.

5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?

Il n'est pas envisagé actuellement d'attribuer systématiquement de DOI à la production d'un jeu de données. Cette attribution reste à la

charge du responsable scientifique de chaque partenaire du projet réalisant des analyses.

Chaque donnée est identifiée par une numérotation particulière défini pour chaque partenaire en charge des prélèvements.

A savoir :

Anses Maisons-Alfort :

Dans le cadre du WP1, afin qu'ils soient faciles à retrouver et identifier, le nom du fichier de chaque rapport comporte la date et le type de réunion et dans le cadre d'une présentation les initiales de la personne qui l'a réalisé (exemple : 2023_01_15_PresentationWP1_OF).

Les données sont référencées par un code défini en interne par chaque partenaire qui permet de les identifier et les retrouver facilement. A titre d'exemple dans le WP1, le nom du fichier de chaque rapport comportera la date et le type de réunion et dans le cadre d'une présentation les initiales de la personne qui l'a réalisé (exemple : 2023_01_15_PresentationWP1_OF).

Les publications produites seront identifiées par un DOI.

Anses Ploufragan (filière volailles) :

Pour le WP2, un N° de série a été attribué au projet: 23TP01-AA (où AA est le N° de l'analyse, 23TP01-01 signifie que c'est la première mise en analyse du projet Clost'Abat). Les échantillons sont numérotés avec le code 23TP01-XX où XX est le N° de l'échantillon. Cette numérotation est celle décrite dans les documents qualité de l'unité HQPAP (PPN/PS/0200).

Hôpital Saint-Antoine :

Dans le cadre du WP4, un numéro interne de suivi de souches est attribué à chaque souche avérée *Clostridioides difficile*. Exemple : CD23-037, avec :

- CD : pour *Clostridioides difficile*
- 23 : les deux derniers chiffres de l'année en cours
- 037 : numéro d'incrémentation des souches du CNR.

Ce numéro suit la souche : tableau Excel, site Pasteur, feuilles de paillasse, fichier .fsa, compte rendu ...

UMR PAM :

Chaque nouvelle expérience est identifiée par un code selon la condition testée. Par exemple, pour les essais d'humidité relative sur les cellules végétatives, le code suivant a été utilisé : LS_CP_HR_CV001

- LS : nom de l'expérimentateur
- CP : nom de l'espèce bactérienne
- HR : humidité relative
- CV : cellules végétatives
- 001 : numéro d'expérience référencé à une date dans le cahier de laboratoire et dans un tableur

Les publications produites seront identifiées par un DOI.

IFIP (filière porcine) :

Pour le WP2, les prélèvements sont identifiés selon une nomenclature standardisée. Le n° d'échantillon est composé du code projet CBAT suivi du code Abattoir Ab, HP ou TL, du n° de campagne 1, 2 ou 3, du numéro de lot prélevé le jour de la campagne, de 1, à 4, du numéro de porc prélevé, de 1, à 5 et du type de prélèvement, F pour fèces, C pour carcasses et P pour le muscle Poitrine : CBAT-HP1-1-4-P.

pour les prélèvements d'environnements, les codes sont comme suit : Nom du projet CBAT, code abattoir, n° de campagne, A pour abattoir et D pour découpe, air 1 à air 3 pour les échantillons d'air, E1 à E5 pour les échantillons de surface: CBAT-HP1-D-E1 ou CBAT-HP1-D-air1.

Pour les souches détectées et stockées, les numéros suivent la même nomenclature que le prélèvement auquel est ajouté col1 à col5 pour colonie 1 à 5 : CBAT-Ab1-3-3-C-col4

IDELE (filière bovine) :

Pour le WP2, les prélèvements de chaque campagne sont identifiés de la même manière, avec un code spécifique pour chaque campagne. Chaque campagne est numéroté de 1 à 8 avec la localisation du prélèvement. Les fèces sont identifiés avec F, les carcasses avec C, les muscles avec M, les environnements abattoir EA, les environnement découpe ED.

Par exemple, l'échantillon de fèces de l'animal 1 de la campagne numéro 1 est noté : C1F1 : C1 (campagne 1, fèces animal numéro 1). Et ainsi de suite ...

6. Responsabilités et ressources en matière de gestion des données

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

La gestion des données (saisie, production, stockage, sauvegarde) est assurée par les responsables scientifiques de chaque partenaire, conformément aux règles de leurs institutions.

Le coordinateur du projet est responsable de la mise en œuvre et du suivi du PGD.

6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

Le financement du stockage local des données est assuré par l'institut hôte, l'Anses qui centralisera l'ensemble des métadonnées générées dans le cadre du projet.

L'accessibilité et la réutilisation des données ouvertes seront assurées par prise de contact avec le coordinateur et les responsables scientifiques de chaque WP, leurs adresses électroniques se trouvant sur le site internet dédié au projet.

Du temps dédié à la gestion des données est prévu, notamment par le coordinateur, afin de garantir :

- la qualité des données
- leur documentation
- leur conformité aux principes FAIR (Findable, Accessible, Interoperable, Reusable)

L'utilisation de formats ouverts et de standards partagés contribue à assurer l'interopérabilité et la réutilisation des données.

WP3-Détection de *C. perfringens* et *C. difficile* dans les échantillons

1. Description des données et collecte ou réutilisation de données existantes

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Les données seront générées à partir de la mise en oeuvre du protocole qui a été mis au point lors du premier semestre du projet et des résultats obtenus à partir de l'analyse des prélèvements collectés dans le WP2.

Le protocole a été mis au point et appliqué par les 3 partenaires. Les résultats de détection sont saisis sur des fichiers en interne, propres à chaque partenaire et sur le fichier partagé sur RESANA.

Les isolats sont expédiés pour confirmation au laboratoire de l'Anses de Maisons-Alfort pour *C. perfringens* et au CNR *C. difficile* à l'hôpital St Antoine pour *C. difficile*.

Les résultats de caractérisation sont déposés sur RESANA.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Les données produites sont :

- résultats de détection (présence/absence)
- données d'isolement (souches)
- conditions expérimentales

2. Documentation et qualité des données

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

Le protocole d'isolement des souches de *C. perfringens* et de *C. difficile* est commun aux 3 filières et a été utilisé. Hormis les

souches isolées, aucune autre donnée n'est associée à ces souches à ce stade.
Un fichier de résultats commun aux 3 filières est disponible sous RESANA

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Pour chaque analyse, un témoin positif (*C. perfringens* et *C. difficile*) est utilisé de manière à valider le protocole d'isolement des souches. Des témoins négatifs sont aussi utilisés pour valider la bonne mise en œuvre des analyses.

Chaque manipulation en anaérobiose est contrôlée avec un indicateur spécifique.

La traçabilité de la manipulation est assurée via un fichier excel partagé entre le laboratoire et le responsable scientifique du projet (réactifs utilisés, matériel, dates, etc.).

Fiches de conformité des milieux fournisseurs utilisés

3. Stockage et sauvegarde pendant le processus de recherche

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Les données et métadonnées seront stockées et sauvegardées tout au long du projet dans des cahiers de laboratoire repertoriés et stockés dans des armoires fermant à clé ou des classeurs, sur les ordinateurs des membres du projet, sur des disques durs externes, sur les serveurs des instituts partenaires du projet et sur un serveur intranet de l'Anses uniquement dédié aux données générées lors du projet et accessible uniquement par les membres du projet (en lecture seule sauf pour les responsables scientifiques de chaque partenaire).

IFIP (filière porcine) :

Enregistrement des échantillons/heure de mesure des sondes déposées : 1 fichier Excel de traça/campagne pour numéroter les échantillons

Traitement des échantillons : sur fichier Excel (différent de l'enregistrement)

Mesure Pendant les campagnes, sur papier puis stockage sur OneDrive (enregistrement des échantillons)

Stockage :

OneDrive dans un datacenter de Microsoft en France

IDELE (filière bovine) :

Le traitement des échantillons et les résultats par campagne sont traités sur un fichier Excel séparé. Enfin, tous les résultats sont saisis et enregistrés sur un fichier partagé RESANA avec une copie sur un fichier Excel en interne.

Stockage :

Les données sont stockées sur le disque dur de l'ordinateur ainsi que sur le réseau interne à l'IDELE (sauvegarde automatique plusieurs fois par jour).

Anses Ploufragan (filière volailles) :

Les informations de traçabilité (enregistrements des échantillons, réactifs utilisés, heure de mise en incubation, résultats de lecture des géloses...) sont reportées sur des fiche dédiées puis sont ensuite enregistrées dans un fichier Excel dédié, enregistré sur le réseau de l'Anses dans un dossier dont l'accès est réservé aux personnes impliquées dans le projet au sein du laboratoire de Ploufragan-Plouzané-Niort. Enfin, tous les résultats sont saisis et enregistrés sur un fichier partagé RESANA .

Stockage :

Les données sont stockées sur le réseau de l'Anses (copie du réseau quotidienne sur un autre serveur).

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

IFIP (filière porcine) :

Stockage sécurisé par double authentification pour l'accès. Les membres du projet peuvent avoir accès par invitation et

inscription dans notre Azur Directory.

IDELE (filière bovine) :

Les données sont protégées par le mot de passe de l'ordinateur de chaque utilisateur. Elles ne sont disponibles que pour les personnes impliquées au projet ClostAbat.

Anses Ploufragan (filière volailles) :

Les données sont stockées sur le réseau de l'Anses avec un accès restreint aux personnes impliquées dans le projet au sein du laboratoire de l'Anses de Ploufragan-Plouzané-Niort.

4. Exigences légales et éthiques, codes de conduite

4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Les données à caractère personnel, lorsqu'elles existent, sont anonymisées conformément à la réglementation en vigueur (RGPD).

4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Les droits de propriété intellectuelle et les modalités de gestion des données sont définis dans l'accord de consortium signé par l'ensemble des partenaires.

Conformément à la loi n°2016-1321 pour une République numérique, les données issues d'une recherche financée sur fonds publics ont vocation à être rendues accessibles, sous réserve de l'absence de contraintes juridiques ou de confidentialité.

4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Le projet ne soulève pas de problématiques éthiques spécifiques.

5. Partage des données et conservation à long terme

5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?

L'ensemble des données sont partagées entre les partenaires du projet via l'espace sécurisé RESANA.

Les prélèvements réalisés lors des campagnes de prélèvement sont détruit après analyse.

Les données concernant le nom des établissements participants resteront confidentielles et une anonymisation des données de séquences humaines sera respectée.

Les données de valorisation (avancements du projet, affiches, conférences, articles ...) seront accessibles progressivement, au fur et à mesure de leur intégration sur le site internet tout au long du projet. Nous souhaitons rendre ces données accessibles le plus rapidement possible afin de rendre visible l'état d'avancement du projet et dans les principes de la science ouverte.

L'ensemble des données collectées non confidentielles dans le cadre de ce projet seront disponibles en libre accès après l'embargo de publication finale des données traitées comme indiqué dans l'accord de consortium.

5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

Pour le moment, les partenaires du projet s'engagent à conserver l'ensemble des données/métadonnées produites sur les serveurs de leurs instituts respectifs pendant toute la durée du projet et à minima pendant 3 années supplémentaires.

Tous les cahiers de laboratoire papier seront conservés et archivés.

Les données génomiques déposées dans Genbank seront préservées dans le cadre des missions du NCBI (<https://www.ncbi.nlm.nih.gov/genbank/>).

Plus particulièrement, pour les partenaires :

Anses Maisons-Alfort :

L'ensemble des souches de *C. perfringens* envoyées par les partenaires seront conservées à -80°C.

Les données collectées sont conservées sur le réseau de l'Anses.

Anses Ploufragan (filiale volailles) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés en DASRI lorsque la mise en collection des isolats est terminée. Les isolats sont stockés au surgélateur. Pour *C. perfringens*, deux tubes sont préparés par isolat : un est stocké à Ploufragan, le second est envoyé à Maisons-Alfort. Pour *C. difficile*, les isolats sont repiqués sur ChromID à partir du tube stocké au surgélateur puis envoyé à l'hôpital Saint-Antoine (WP3).

Les prélèvements destinés à l'étude des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant extraction et séquençage (WP6).

les données de séquençage seront déposées dans des bases publiques (SRA).

Les données collectées sont conservées sur le réseau de l'Anses.

Institut Pasteur :

Les données de séquençage des souches de *C. perfringens* et *C. difficile* sont stockées sur le réseau de l'Institut Pasteur.

Hôpital Saint-Antoine :

Archivage papier (feuilles de paillasse + CR) dans le bureau du CNR *Clostridioides difficile* au sein de l'Hôpital St Antoine.

Archivage informatique (réseau APHP + disque dur externe pour les fichiers .fsa + Résana + site Pasteur)

UMR PAM :

L'ensemble des données produites seront conservées sur le long terme par un cahier de laboratoire et un disque dur externe. Le cahier de laboratoire et le disque dur externe seront conservés à vie dans des armoires fermant à clé dans l'unité de recherche.

IDELE (filiale bovine) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés lorsque la confirmation des souches isolées (*C. perfringens* par ANSES et *C. difficile* par Hôpital Saint-Antoine) a été réalisée. Les souches isolées sont conservées à -80°C en parallèle de l'envoi à l'ANSES Maisons-Alfort et à l'Hôpital Saint-Antoine (WP3). Les prélèvements destinés à l'étude des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant envoi à l'ANSES de Ploufragan (WP6).

Pour le WP3, les souches potentielles issues de la filière bovine sont identifiées de la manière suivante (commune avec les autres partenaires) : CBAT-P-B-XXX (ClostAbat - Perfringens - Bovin - numéro de 1 à XXX) et CBAT-D-B-XXX (ClostAbat - Difficile - Bovin - numéro de 1 à XXX).

IFIP (filiale porcine) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés après traitement lors de chaque campagne. Les souches isolées sont conservées à -80°C en parallèle de l'envoi à l'ANSES et à l'hôpital Saint-Antoine (WP3).

Les prélèvements destinés à l'étude des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant envoi à l'ANSES Ploufragan (WP6).

Les données obtenues sont conservées dans l'espace de stockage des données de l'IFIP par archivage

5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

Les outils logiciels pour accéder et utiliser les données/métadonnées du projet sont standards et ouverts pour faciliter le partage et la réutilisation à long terme des données.

Néanmoins, certains logiciels utilisés pour les analyses sont restreints d'accès :

- GeneMapper : logiciel installé sur les ordinateurs du CNR *Clostridioides difficile*. Code de connexion (login + mot de passe) unique pour l'équipe du CNR

- Webribo pour la lecture des fichiers .fsa (données brutes) : site internet. Chaque nouvel utilisateur doit se créer son compte.

Pas de visualisation possible des données des autres comptes (mais possibilité de retélécharger le fichier .fsa de la souche issu du séquenceur).

- BioNumerics : logiciel d'analyse des profils de typage moléculaire M13 pour les souches de *C. perfringens*.

5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?

Il n'est pas envisagé actuellement d'attribuer systématiquement de DOI à la production d'un jeu de données. Cette attribution reste à la

charge du responsable scientifique de chaque partenaire du projet réalisant des analyses.

Chaque donnée est identifiée par une numérotation particulière défini pour chaque partenaire en charge des prélèvements.

A savoir :

Anses Maisons-Alfort :

Dans le cadre du WP1, afin qu'ils soient faciles à retrouver et identifier, le nom du fichier de chaque rapport comporte la date et le type de réunion et dans le cadre d'une présentation les initiales de la personne qui l'a réalisé (exemple : 2023_01_15_PresentationWP1_OF).

Les données sont référencées par un code défini en interne par chaque partenaire qui permet de les identifier et les retrouver facilement. A titre d'exemple dans le WP1, le nom du fichier de chaque rapport comportera la date et le type de réunion et dans le cadre d'une présentation les initiales de la personne qui l'a réalisé (exemple : 2023_01_15_PresentationWP1_OF).

Les publications produites seront identifiées par un DOI.

Anses Ploufragan (filière volailles) :

Pour le WP2, un N° de série a été attribué au projet: 23TP01-AA (où AA est le N° de l'analyse, 23TP01-01 signifie que c'est la première mise en analyse du projet Clost'Abat). Les échantillons sont numérotés avec le code 23TP01-XX où XX est le N° de l'échantillon. Cette numérotation est celle décrite dans les documents qualité de l'unité HQPAP (PPN/PS/0200).

Hôpital Saint-Antoine :

Dans le cadre du WP4, un numéro interne de suivi de souches est attribué à chaque souche avérée *Clostridioides difficile*.
Exemple : CD23-037, avec :

- CD : pour *Clostridioides difficile*
- 23 : les deux derniers chiffres de l'année en cours
- 037 : numéro d'incrémentation des souches du CNR.

Ce numéro suit la souche : tableau Excel, site Pasteur, feuilles de paillasse, fichier .fsa, compte rendu ...

UMR PAM :

Chaque nouvelle expérience est identifiée par un code selon la condition testée. Par exemple, pour les essais d'humidité relative sur les cellules végétatives, le code suivant a été utilisé : LS_CP_HR_CV001

- LS : nom de l'expérimentateur
- CP : nom de l'espèce bactérienne
- HR : humidité relative
- CV : cellules végétatives
- 001 : numéro d'expérience référencé à une date dans le cahier de laboratoire et dans un tableur

Les publications produites seront identifiées par un DOI.

IFIP (filière porcine) :

Pour le WP2, les prélèvements sont identifiés selon une nomenclature standardisée. Le n° d'échantillon est composé du code projet CBAT suivi du code Abattoir Ab, HP ou TL, du n° de campagne 1, 2 ou 3, du numéro de lot prélevé le jour de la campagne, de 1, à 4, du numéro de porc prélevé, de 1, à 5 et du type de prélèvement, F pour fèces, C pour carcasses et P pour le muscle Poitrine : CBAT-HP1-1-4-P.

pour les prélèvements d'environnements, les codes sont comme suit : Nom du projet CBAT, code abattoir, n° de campagne, A pour abattoir et D pour découpe, air 1 à air 3 pour les échantillons d'air, E1 à E5 pour les échantillons de surface: CBAT-HP1-D-E1 ou CBAT-HP1-D-air1.

Pour les souches détectées et stockées, les numéros suivent la même nomenclature que le prélèvement auquel est ajouté col1 à col5 pour colonie 1 à 5 : CBAT-Ab1-3-3-C-col4

IDELE (filière bovine) :

Pour le WP2, les prélèvements de chaque campagne sont identifiés de la même manière, avec un code spécifique pour chaque campagne. Chaque campagne est numéroté de 1 à 8 avec la localisation du prélèvement. Les fèces sont identifiés avec F, les carcasses avec C, les muscles avec M, les environnements abattoir EA, les environnement découpe ED.

Par exemple, l'échantillon de fèces de l'animal 1 de la campagne numéro 1 est noté : C1F1 : C1 (campagne 1, fèces animal numéro 1). Et ainsi de suite ...

6. Responsabilités et ressources en matière de gestion des données

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

La gestion des données (saisie, production, stockage, sauvegarde) est assurée par les responsables scientifiques de chaque partenaire, conformément aux règles de leurs institutions.

Le coordinateur du projet est responsable de la mise en œuvre et du suivi du PGD.

6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

Le financement du stockage local des données est assuré par l'institut hôte, l'Anses qui centralisera l'ensemble des métadonnées générées dans le cadre du projet.

L'accessibilité et la réutilisation des données ouvertes seront assurées par prise de contact avec le coordinateur et les responsables scientifiques de chaque WP, leurs adresses électroniques se trouvant sur le site internet dédié au projet.

Du temps dédié à la gestion des données est prévu, notamment par le coordinateur, afin de garantir :

- la qualité des données
- leur documentation
- leur conformité aux principes FAIR (Findable, Accessible, Interoperable, Reusable)

L'utilisation de formats ouverts et de standards partagés contribue à assurer l'interopérabilité et la réutilisation des données.

WP4-Diversité génétique et toxinotype des souches de *C. perfringens* et *C. difficile*

1. Description des données et collecte ou réutilisation de données existantes

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Les données du WP4 sont générées à partir des souches isolées dans le WP3.

Les analyses comprennent le typage moléculaire (PCR-M13, ribotypage), le toxinotype et les profils d'antibiorésistance des souches.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Les données produites sont :

- résultats de toxinotype (présence des gènes codant les toxines)
- résultats de typage moléculaire (PCR M13 pour *C. perfringens* et ribotypage pour *C. difficile*)
- résultats de profils d'antibiorésistance (*C. difficile*)

Formats :

- XLSX, CSV, JPE, PDF
- FSA (séquenceurs ABI 3500 Genetic Analyzer)

Pour *C. difficile*, les résultats de ribotypage au format .fsa sont transférés à l'aide d'une clé USB pour analyse sur le site en ligne *webribo*. L'ensemble des données est ensuite enregistré sur le site *epidemio.pasteur.fr* (©voozanoo/epiconcept 2006).

2. Documentation et qualité des données

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

Les méthodologies utilisées pour la caractérisation des souches de *C. perfringens* et de *C. difficile* sont validées chez chacun des partenaires en charge de cette tâche.

A savoir :

Ansès Maisons-Alfort (souches de *C. perfringens*) :

- Ensemencement une colonie sur gélose COS
- Extraction ADN
- PCR Toxinotype
- PCR Typage moléculaire M13
- Run sur Roche 480 Analyser
- Analyse du run et rendu de résultat
- Analyse des profils de typage moléculaire par le logiciel BioNumerics (Applied Maths)

CNR *C.difficile* (souches de *C. difficile*) :

- MALDI-TOF
- Enregistrement des informations et saisie des résultats sur *epidemio.pasteur.fr*
- Ensemencement une colonie sur gélose CLO
- Extraction ADN
- PCR Multiplex
- PCR Ribotype
- Run sur ABI 3500 Genetic Analyzer
- Analyse du run et rendu de résultat
- Antibiogramme et lecture

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Méthodes validés pour chaque partenaire

Formation des opérateurs

Contrôles internes lors du traitement des échantillons

Fiche de suivi et traçabilité des échantillons (n°lot, réactif, fournisseur, suivi de température, matériels utilisés ...)

Vérification et étalonnage des appareils de mesure

Procédures qualité internes

3. Stockage et sauvegarde pendant le processus de recherche

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Anses Maisons-Alfort (souches de *C. perfringens*) :

Les informations de traçabilité (enregistrements des souches, réactifs utilisés, réalisation des analyses, résultats...) sont reportées sur des fiche dédiées puis sont ensuite enregistrées dans le fichier Excel dédié, enregistré sur le réseau de l'Anses dans un dossier dont l'accès est réservé aux personnes impliquées dans le projet au sein du laboratoire de sécurité des aliments. Enfin, tous les résultats sont saisis et enregistrés sur un fichier partagé RESANA .

Stockage :

Les données sont stockées sur le réseau de l'Anses (copie du réseau quotidienne sur un autre serveur).

Hôpital Saint-Antoine (souches de *C. difficile*) :

MALDI : Le fichier papier de résultats est gardé dans le dossier de l'étude. Les résultats sont également reportés dans le tableau Excel contenant les données des souches ainsi que les résultats.

Pour les souches identifiées au MALDI TOF comme étant des *Clostridioïdes difficile* :

- une feuille de suivi de la souche est créée et remplie avec les résultats (feuille du CNR, support papier archivé)
- un enregistrement des souches *Clostridioïdes difficile* est fait sur epidemiopasteur.fr (©voozanoo/epiconcept 2006)
- Les données du run sur ABI 3500 Genetic Analyzer sont conservés sur le PC de l'automate. Ces données sont copiées/collées sur une clé USB destinée uniquement au transfert entre le PC de l'automate et le PC APHP du technicien. Ces données sont copiées/collées dans un dossier sur le réseau protégé APHP. Elles sont dupliquées de façon régulière sur un disque dur externe. Ces données sont transférées pour analyse dans Genemapper (logiciel d'analyse des facteurs de virulence) ainsi que sur le site Webribo (identification du Ribotype) (Webribo.ages.at). Les résultats des interprétations des PCR Multiplex et Ribotype sont notés sur chaque feuille de suivi de souches et sont saisis sur le tableau Excel contenant les données des souches ainsi que les résultats.
- Les résultats des antibiogrammes sont inscrits sur la feuille de suivi de chaque souche et saisis sur le tableau Excel contenant les données des souches.
- Quand tout a été validé définitivement au sein du laboratoire CNR *C.diff*, le tableau Excel contenant les données de base et les résultats obtenus sont transmis sur RESANA (plateforme collaborative de l'État pour vos groupes de travail.)

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

Anses Maisons-Alfort (souches de *C. perfringens*) :

Les données sont stockées sur le réseau de l'Anses avec un accès restreint aux personnes impliquées dans le projet au sein du laboratoire de sécurité des aliments de l'Anses.

Hôpital Saint-Antoine (souches de *C. difficile*) :

Pas de sortie des données hors laboratoire (pas d'ordinateur personnel, pas de télétravail, pas de données sur clés, pas de dépôt des données sur cloud autre que Résana, pas de sortie des feuilles hors du laboratoire...)

Seules les personnes habilitées à se connecter au serveur de Bactériologie de l'hôpital St Antoine ont accès aux données stockées par le CNR *Clostridioïdes difficile*. La sécurité informatique du réseau est assurée par le service informatique de l'hôpital St Antoine

Archivage des feuilles de résultats (feuille de paillasse et CR de résultats) au sein du laboratoire (pas de feuilles volantes ni de destruction des feuilles de résultats).

4. Exigences légales et éthiques, codes de conduite

4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Les données à caractère personnel, lorsqu'elles existent, sont anonymisées conformément à la réglementation en vigueur (RGPD).

4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les

données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Les droits de propriété intellectuelle et les modalités de gestion des données sont définis dans l'accord de consortium signé par l'ensemble des partenaires.

Conformément à la loi n°2016-1321 pour une République numérique, les données issues d'une recherche financée sur fonds publics ont vocation à être rendues accessibles, sous réserve de l'absence de contraintes juridiques ou de confidentialité.

4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Le projet ne soulève pas de problématiques éthiques spécifiques.

5. Partage des données et conservation à long terme

5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?

L'ensemble des données sont partagées entre les partenaires du projet via l'espace sécurisé RESANA.

Les prélèvements réalisés lors des campagnes de prélèvement sont détruit après analyse.

Les données concernant le nom des établissements participants resteront confidentielles et une anonymisation des données de séquences humaines sera respectée.

Les données de valorisation (avancements du projet, affiches, conférences, articles ...) seront accessibles progressivement, au fur et à mesure de leur intégration sur le site internet tout au long du projet. Nous souhaitons rendre ces données accessibles le plus rapidement possible afin de rendre visible l'état d'avancement du projet et dans les principes de la science ouverte.

L'ensemble des données collectées non confidentielles dans le cadre de ce projet seront disponibles en libre accès après l'embargo de publication finale des données traitées comme indiqué dans l'accord de consortium.

5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

Pour le moment, les partenaires du projet s'engagent à conserver l'ensemble des données/métadonnées produites sur les serveurs de leurs instituts respectifs pendant toute la durée du projet et à minima pendant 3 années supplémentaires.

Tous les cahiers de laboratoire papier seront conservés et archivés.

Les données génomiques déposées dans Genbank seront préservées dans le cadre des missions du NCBI (<https://www.ncbi.nlm.nih.gov/genbank/>).

Plus particulièrement, pour les partenaires :

Anses Maisons-Alfort :

L'ensemble des souches de *C. perfringens* envoyées par les partenaires seront conservées à -80°C.

Les données collectées sont conservées sur le réseau de l'Anses.

Anses Ploufragan (filière volailles) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés en DASRI lorsque la mise en collection des isolats est terminée. Les isolats sont stockés au surgélateur. Pour *C. perfringens*, deux tubes sont préparés par isolat : un est stocké à Ploufragan, le second est envoyé à Maisons-Alfort. Pour *C. difficile*, les isolats sont repiqués sur ChromID à partir du tube stocké au surgélateur puis envoyé à l'hôpital Saint-Antoine (WP3).

Les prélèvements destinés à l'étude des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant extraction et séquençage (WP6).

les données de séquençage seront déposées dans des bases publiques (SRA).

Les données collectées sont conservées sur le réseau de l'Anses.

Institut Pasteur :

Les données de séquençage des souches de *C. perfringens* et *C. difficile* sont stockées sur le réseau de l'Institut Pasteur.

Hôpital Saint-Antoine :

Archivage papier (feuilles de paillasse + CR) dans le bureau du CNR *Clostridioides difficile* au sein de l'Hôpital St Antoine.

Archivage informatique (réseau APHP + disque dur externe pour les fichiers .fsa + Résana + site Pasteur)

UMR PAM :

L'ensemble des données produites seront conservées sur le long terme par un cahier de laboratoire et un disque dur externe. Le cahier de laboratoire et le disque dur externe seront conservés à vie dans des armoires fermant à clé dans l'unité de recherche.

IDELE (filière bovine) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés lorsque la confirmation des souches isolées (*C. perfringens* par ANSES et *C. difficile* par Hôpital Saint-Antoine) a été réalisée. Les souches isolées sont conservées à -80°C en parallèle de l'envoi à l'ANSES Maisons-Alfort et à l'Hôpital Saint-Antoine (WP3). Les prélèvements destinés à l'étude des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant envoi à l'ANSES de Ploufragan (WP6).

Pour le WP3, les souches potentielles issues de la filière bovine sont identifiées de la manière suivante (commune avec les autres partenaires) : CBAT-P-B-XXX (ClostAbat - Perfringens - Bovin - numéro de 1 à XXX) et CBAT-D-B-XXX (ClostAbat - Difficile - Bovin - numéro de 1 à XXX).

IFIP (filière porcine) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés après traitement lors de chaque campagne. Les souches isolées sont conservées à -80°C en parallèle de l'envoi à l'ANSES et à l'hôpital Saint-Antoine (WP3).

Les prélèvements destinés à l'étude des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant envoi à l'ANSES Ploufragan (WP6).

Les données obtenues sont conservées dans l'espace de stockage des données de l'IFIP par archivage

5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

Les outils logiciels pour accéder et utiliser les données/métadonnées du projet sont standards et ouverts pour faciliter le partage et la réutilisation à long terme des données.

Néanmoins, certains logiciels utilisés pour les analyses sont restreints d'accès :

- GeneMapper : logiciel installé sur les ordinateurs du CNR *Clostridioides difficile*. Code de connexion (login + mot de passe) unique pour l'équipe du CNR
- Webribo pour la lecture des fichiers .fsa (données brutes) : site internet. Chaque nouvel utilisateur doit se créer son compte. Pas de visualisation possible des données des autres comptes (mais possibilité de retélécharger le fichier .fsa de la souche issue du séquenceur).
- BioNumerics : logiciel d'analyse des profils de typage moléculaire M13 pour les souches de *C. perfringens*.

5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?

Il n'est pas envisagé actuellement d'attribuer systématiquement de DOI à la production d'un jeu de données. Cette attribution reste à la charge du responsable scientifique de chaque partenaire du projet réalisant des analyses.

Chaque donnée est identifiée par une numérotation particulière définie pour chaque partenaire en charge des prélèvements.

A savoir :

Anses Maisons-Alfort :

Dans le cadre du WP1, afin qu'ils soient faciles à retrouver et identifier, le nom du fichier de chaque rapport comporte la date et le type de réunion et dans le cadre d'une présentation les initiales de la personne qui l'a réalisé (exemple : 2023_01_15_PresentationWP1_OF).

Les données sont référencées par un code défini en interne par chaque partenaire qui permet de les identifier et les retrouver facilement. A titre d'exemple dans le WP1, le nom du fichier de chaque rapport comportera la date et le type de réunion et dans le cadre d'une présentation les initiales de la personne qui l'a réalisé (exemple : 2023_01_15_PresentationWP1_OF).

Les publications produites seront identifiées par un DOI.

Anses Ploufragan (filière volailles) :

Pour le WP2, un N° de série a été attribué au projet: 23TP01-AA (où AA est le N° de l'analyse, 23TP01-01 signifie que c'est la première mise en analyse du projet Clost'Abat). Les échantillons sont numérotés avec le code 23TP01-XX où XX est le N° de

l'échantillon. Cette numérotation est celle décrite dans les documents qualité de l'unité HQPAP (PPN/PS/0200).

Hôpital Saint-Antoine :

Dans le cadre du WP4, un numéro interne de suivi de souches est attribué à chaque souche avérée *Clostridioides difficile*.
Exemple : CD23-037, avec :

- CD : pour *Clostridioides difficile*
- 23 : les deux derniers chiffres de l'année en cours
- 037 : numéro d'incrémentation des souches du CNR.

Ce numéro suit la souche : tableau Excel, site Pasteur, feuilles de paillasse, fichier .fsa, compte rendu ...

UMR PAM :

Chaque nouvelle expérience est identifiée par un code selon la condition testée. Par exemple, pour les essais d'humidité relative sur les cellules végétatives, le code suivant a été utilisé : LS_CP_HR_CV001

- LS : nom de l'expérimentateur
- CP : nom de l'espèce bactérienne
- HR : humidité relative
- CV : cellules végétatives
- 001 : numéro d'expérience référencé à une date dans le cahier de laboratoire et dans un tableur

Les publications produites seront identifiées par un DOI.

IFIP (filiale porcine) :

Pour le WP2, les prélèvements sont identifiés selon une nomenclature standardisée. Le n° d'échantillon est composé du code projet CBAT suivi du code Abattoir Ab, HP ou TL, du n° de campagne 1, 2 ou 3, du numéro de lot prélevé le jour de la campagne, de 1, à 4, du numéro de porc prélevé, de 1, à 5 et du type de prélèvement, F pour fèces, C pour carcasses et P pour le muscle Poitrine : CBAT-HP1-1-4-P.

pour les prélèvements d'environnements, les codes sont comme suit : Nom du projet CBAT, code abattoir, n° de campagne, A pour abattoir et D pour découpe, air 1 à air 3 pour les échantillons d'air, E1 à E5 pour les échantillons de surface: CBAT-HP1-D-E1 ou CBAT-HP1-D-air1.

Pour les souches détectées et stockées, les numéros suivent la même nomenclature que le prélèvement auquel est ajouté col1 à col5 pour colonie 1 à 5 : CBAT-Ab1-3-3-C-col4

IDELE (filiale bovine) :

Pour le WP2, les prélèvements de chaque campagne sont identifiés de la même manière, avec un code spécifique pour chaque campagne. Chaque campagne est numéroté de 1 à 8 avec la localisation du prélèvement. Les fèces sont identifiés avec F, les carcasses avec C, les muscles avec M, les environnements abattoir EA, les environnements découpe ED.

Par exemple, l'échantillon de fèces de l'animal 1 de la campagne numéro 1 est noté : C1F1 : C1 (campagne 1, fèces animal numéro 1). Et ainsi de suite ...

6. Responsabilités et ressources en matière de gestion des données

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

La gestion des données (saisie, production, stockage, sauvegarde) est assurée par les responsables scientifiques de chaque partenaire, conformément aux règles de leurs institutions.

Le coordinateur du projet est responsable de la mise en œuvre et du suivi du PGD.

6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

Le financement du stockage local des données est assuré par l'institut hôte, l'Anses qui centralisera l'ensemble des métadonnées générées dans le cadre du projet.

L'accessibilité et la réutilisation des données ouvertes seront assurées par prise de contact avec le coordinateur et les responsables scientifiques de chaque WP, leurs adresses électroniques se trouvant sur le site internet dédié au projet.

Du temps dédié à la gestion des données est prévu, notamment par le coordinateur, afin de garantir :

- la qualité des données
- leur documentation
- leur conformité aux principes FAIR (Findable, Accessible, Interoperable, Reusable)

L'utilisation de formats ouverts et de standards partagés contribue à assurer l'interopérabilité et la réutilisation des données.

WP5-Structure des populations et attribution de source

1. Description des données et collecte ou réutilisation de données existantes

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Les données du WP5 sont produites à partir des souches isolées et caractérisées dans les WP précédents, ainsi que de données de prélèvements environnementaux collectées en abattoirs.

Elles reposent sur la génération et l'exploitation de données génomiques et métagénomiques afin d'analyser la structure des populations de *Clostridium perfringens* et *Clostridioides difficile* et d'attribuer leur origine aux différentes filières et environnements.

Un ensemble d'isolats représentatifs des filières animales, de l'environnement et de cas humains est séquencé, complété par des génomes déjà disponibles chez certains partenaires et dans des bases de données publiques. Les séquences sont assemblées, analysées et comparées afin de définir la structure des populations, d'identifier des marqueurs génomiques associés à l'origine ou à la persistance des souches et de développer des outils d'attribution de source.

En parallèle, la diversité et la composition des communautés bactériennes environnementales présentes en abattoirs sont étudiées par métabarcoding 16S afin d'explorer les liens entre microbiote, environnement et présence des deux pathogènes.

L'ensemble de ces données est exploité à l'aide d'approches statistiques, phylogénétiques, de modélisation d'attribution de source et, de manière exploratoire, de méthodes d'apprentissage automatique. Aucune donnée personnelle n'est collectée dans ce WP.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Les données produites comprennent :

- données de séquençage brutes : FASTQ,
- assemblages de génomes : FASTA,
- données d'annotation et de pangenome : CSV/XLSX,
- matrices cgMLST, SNP, MLST : CSV/XLSX,
- arbres phylogénétiques : formats Newick / PDF,
- données métagénomiques 16S : FASTQ + tables d'abondance.

Les volumes de données sont importants pour les séquences mais compatibles avec les capacités de stockage institutionnelles.

2. Documentation et qualité des données

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

Le code échantillon correspond à celui de l'échantillon de départ et les métadonnées correspondantes seront rapatriées des WP2 et 3 (type d'abattoir, échantillon prélevé, date de prélèvement, résultat de détection pour *C. difficile* et *C. perfringens*...).

L'extraction d'ADN est faite selon un protocole interne à l'unité HQPAP. L'amplification des régions V3-V4 de l'ADNr 16S ainsi que le séquençage en Illumina sont gérées selon les protocoles de la plateforme de séquençage du laboratoire de Ploufragan-Plouzané-Niort.

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Les contrôles qualité incluent :

- évaluations qualité des lectures de séquençage,
- contrôles d'assemblage,
- validation bioinformatique des pangenomes,
- contrôles statistiques et comparatifs pour les analyses métagénomiques.

3. Stockage et sauvegarde pendant le processus de recherche

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Les données sont stockées :

- sur les serveurs sécurisés des partenaires producteurs,
- sur l'espace collaboratif sécurisé RESANA pour le partage inter-partenaires,
- sur les infrastructures institutionnelles pour l'archivage.

Les sauvegardes sont assurées automatiquement par les services informatiques des institutions.

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

L'accès aux données est restreint aux membres autorisés du projet.

Les données ne contiennent pas de données à caractère personnel.

Les informations permettant l'identification des établissements sont anonymisées.

4. Exigences légales et éthiques, codes de conduite

4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Les données à caractère personnel, lorsqu'elles existent, sont anonymisées conformément à la réglementation en vigueur (RGPD).

4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Les droits de propriété intellectuelle et les modalités de gestion des données sont définis dans l'accord de consortium signé par l'ensemble des partenaires.

Conformément à la loi n°2016-1321 pour une République numérique, les données issues d'une recherche financée sur fonds publics ont vocation à être rendues accessibles, sous réserve de l'absence de contraintes juridiques ou de confidentialité.

4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Le projet ne soulève pas de problématiques éthiques spécifiques.

5. Partage des données et conservation à long terme

5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?

L'ensemble des données sont partagées entre les partenaires du projet via l'espace sécurisé RESANA.

Les prélèvements réalisés lors des campagnes de prélèvement sont détruit après analyse.

Les données concernant le nom des établissements participants resteront confidentielles et une anonymisation des données de séquences humaines sera respectée.

Les données de valorisation (avancements du projet, affiches, conférences, articles ...) seront accessibles progressivement, au fur et à mesure de leur intégration sur le site internet tout au long du projet. Nous souhaitons rendre ces données accessibles le plus rapidement possible afin de rendre visible l'état d'avancement du projet et dans les principes de la science ouverte.

L'ensemble des données collectées non confidentielles dans le cadre de ce projet seront disponibles en libre accès après l'embargo de publication finale des données traitées comme indiqué dans l'accord de consortium.

5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

Pour le moment, les partenaires du projet s'engagent à conserver l'ensemble des données/métadonnées produites sur les serveurs de leurs instituts respectifs pendant toute la durée du projet et à minima pendant 3 années supplémentaires.

Tous les cahiers de laboratoire papier seront conservés et archivés.

Les données génomiques déposées dans Genbank seront préservées dans le cadre des missions du NCBI (<https://www.ncbi.nlm.nih.gov/genbank/>).

Plus particulièrement, pour les partenaires :

Anses Maisons-Alfort :

L'ensemble des souches de *C. perfringens* envoyées par les partenaires seront conservées à -80°C.

Les données collectées sont conservées sur le réseau de l'Anses.

Anses Ploufragan (filière volailles) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés en DASRI lorsque la mise en collection des isolats est terminée. Les isolats sont stockés au surgélateur. Pour *C. perfringens*, deux tubes sont préparés par isolat : un est stocké à Ploufragan, le second est envoyé à Maisons-Alfort. Pour *C. difficile*, les isolats sont repiqués sur ChromID à partir du tube stocké au surgélateur puis envoyé à l'hôpital Saint-Antoine (WP3).

Les prélèvements destinés à l'étude des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant extraction et séquençage (WP6).

Les données de séquençage seront déposées dans des bases publiques (SRA).

Les données collectées sont conservées sur le réseau de l'Anses.

Institut Pasteur :

Les données de séquençage des souches de *C. perfringens* et *C. difficile* sont stockées sur le réseau de l'Institut Pasteur.

Hôpital Saint-Antoine :

Archivage papier (feuilles de paillasse + CR) dans le bureau du CNR *Clostridioides difficile* au sein de l'Hôpital St Antoine.

Archivage informatique (réseau APHP + disque dur externe pour les fichiers .fsa + Résana + site Pasteur)

UMR PAM :

L'ensemble des données produites seront conservées sur le long terme par un cahier de laboratoire et un disque dur externe.

Le cahier de laboratoire et le disque dur externe seront conservés à vie dans des armoires fermant à clé dans l'unité de recherche.

IDELE (filière bovine) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés lorsque la confirmation des souches isolées (*C. perfringens* par ANSES et *C. difficile* par Hôpital Saint-Antoine) a été réalisée. Les souches isolées sont conservées à -80°C en parallèle de l'envoi à l'ANSES Maisons-Alfort et à l'Hôpital Saint-Antoine (WP3). Les prélèvements destinés à l'étude des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant envoi à l'ANSES de Ploufragan (WP6).

Pour le WP3, les souches potentielles issues de la filière bovine sont identifiées de la manière suivante (commune avec les autres partenaires) : CBAT-P-B-XXX (ClostAbat - Perfringens - Bovin - numéro de 1 à XXX) et CBAT-D-B-XXX (ClostAbat - Difficile - Bovin - numéro de 1 à XXX).

IFIP (filière porcine) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés après traitement lors de chaque campagne. Les souches isolées sont conservées à -80°C en parallèle de l'envoi à l'ANSES et à l'hôpital Saint-Antoine (WP3).

Les prélèvements destinés à l'étude des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant envoi à l'ANSES Ploufragan (WP6).

Les données obtenues sont conservées dans l'espace de stockage des données de l'IFIP par archivage

5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

Les outils logiciels pour accéder et utiliser les données/métadonnées du projet sont standards et ouverts pour faciliter le partage et la réutilisation à long terme des données.

Néanmoins, certains logiciels utilisés pour les analyses sont restreints d'accès :

- GeneMapper : logiciel installé sur les ordinateurs du CNR *Clostridioides difficile*. Code de connexion (login + mot de passe) unique pour l'équipe du CNR

- Webribo pour la lecture des fichiers .fsa (données brutes) : site internet. Chaque nouvel utilisateur doit se créer son compte. Pas de visualisation possible des données des autres comptes (mais possibilité de retélécharger le fichier .fsa de la souche issue du séquenceur).

- BioNumerics : logiciel d'analyse des profils de typage moléculaire M13 pour les souches de *C. perfringens*.

5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?

Il n'est pas envisagé actuellement d'attribuer systématiquement de DOI à la production d'un jeu de données. Cette attribution reste à la

charge du responsable scientifique de chaque partenaire du projet réalisant des analyses.

Chaque donnée est identifiée par une numérotation particulière définie pour chaque partenaire en charge des prélèvements.

A savoir :

Anses Maisons-Alfort :

Dans le cadre du WP1, afin qu'ils soient faciles à retrouver et identifier, le nom du fichier de chaque rapport comporte la date et le type de réunion et dans le cadre d'une présentation les initiales de la personne qui l'a réalisé (exemple :

2023_01_15_PresentationWP1_OF).

Les données sont référencées par un code défini en interne par chaque partenaire qui permet de les identifier et les retrouver facilement. A titre d'exemple dans le WP1, le nom du fichier de chaque rapport comportera la date et le type de réunion et dans le cadre d'une présentation les initiales de la personne qui l'a réalisé (exemple : 2023_01_15_PresentationWP1_OF).

Les publications produites seront identifiées par un DOI.

Anses Ploufragan (filière volailles) :

Pour le WP2, un N° de série a été attribué au projet: 23TP01-AA (où AA est le N° de l'analyse, 23TP01-01 signifie que c'est la première mise en analyse du projet Clost'Abat). Les échantillons sont numérotés avec le code 23TP01-XX où XX est le N° de l'échantillon. Cette numérotation est celle décrite dans les documents qualité de l'unité HQPAP (PPN/PS/0200).

Hôpital Saint-Antoine :

Dans le cadre du WP4, un numéro interne de suivi de souches est attribué à chaque souche avérée *Clostridioides difficile*.
Exemple : CD23-037, avec :

- CD : pour *Clostridioides difficile*
- 23 : les deux derniers chiffres de l'année en cours
- 037 : numéro d'incrémentation des souches du CNR.

Ce numéro suit la souche : tableau Excel, site Pasteur, feuilles de paillasse, fichier .fsa, compte rendu ...

UMR PAM :

Chaque nouvelle expérience est identifiée par un code selon la condition testée. Par exemple, pour les essais d'humidité relative sur les cellules végétatives, le code suivant a été utilisé : LS_CP_HR_CV001

- LS : nom de l'expérimentateur
- CP : nom de l'espèce bactérienne
- HR : humidité relative
- CV : cellules végétatives
- 001 : numéro d'expérience référencé à une date dans le cahier de laboratoire et dans un tableur

Les publications produites seront identifiées par un DOI.

IFIP (filière porcine) :

Pour le WP2, les prélèvements sont identifiés selon une nomenclature standardisée. Le n° d'échantillon est composé du code projet CBAT suivi du code Abattoir Ab, HP ou TL, du n° de campagne 1, 2 ou 3, du numéro de lot prélevé le jour de la campagne, de 1, à 4, du numéro de porc prélevé, de 1, à 5 et du type de prélèvement, F pour fèces, C pour carcasses et P pour le muscle Poitrine : CBAT-HP1-1-4-P.

pour les prélèvements d'environnements, les codes sont comme suit : Nom du projet CBAT, code abattoir, n° de campagne, A pour abattoir et D pour découpe, air 1 à air 3 pour les échantillons d'air, E1 à E5 pour les échantillons de surface: CBAT-HP1-D-E1 ou CBAT-HP1-D-air1.

Pour les souches détectées et stockées, les numéros suivent la même nomenclature que le prélèvement auquel est ajouté col1 à col5 pour colonie 1 à 5 : CBAT-Ab1-3-3-C-col4

IDELE (filière bovine) :

Pour le WP2, les prélèvements de chaque campagne sont identifiés de la même manière, avec un code spécifique pour chaque campagne. Chaque campagne est numérotée de 1 à 8 avec la localisation du prélèvement. Les fèces sont identifiés avec F, les carcasses avec C, les muscles avec M, les environnements abattoir EA, les environnements découpe ED.

Par exemple, l'échantillon de fèces de l'animal 1 de la campagne numéro 1 est noté : C1F1 : C1 (campagne 1, fèces animal numéro 1). Et ainsi de suite ...

6. Responsabilités et ressources en matière de gestion des données

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

La gestion des données (saisie, production, stockage, sauvegarde) est assurée par les responsables scientifiques de chaque partenaire, conformément aux règles de leurs institutions.

Le coordinateur du projet est responsable de la mise en œuvre et du suivi du PGD.

6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

Le financement du stockage local des données est assuré par l'institut hôte, l'Anses qui centralisera l'ensemble des métadonnées générées dans le cadre du projet.

L'accessibilité et la réutilisation des données ouvertes seront assurées par prise de contact avec le coordinateur et les responsables scientifiques de chaque WP, leurs adresses électroniques se trouvant sur le site internet dédié au projet.

Du temps dédié à la gestion des données est prévu, notamment par le coordinateur, afin de garantir :

- la qualité des données

- leur documentation
- leur conformité aux principes FAIR (Findable, Accessible, Interoperable, Reusable)

L'utilisation de formats ouverts et de standards partagés contribue à assurer l'interopérabilité et la réutilisation des données.

WP6-Evaluation et proposition de mesures de maîtrise

1. Description des données et collecte ou réutilisation de données existantes

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Les données du WP6 sont produites à partir de la combinaison de données de terrain, de données existantes et de données expérimentales nouvelles, dans le but d'identifier les pratiques à risque en abattoirs et d'évaluer des mesures de maîtrise applicables à *Clostridium perfringens* et *Clostridioides difficile*.

Ce WP s'appuie sur :

- les données environnementales et opérationnelles collectées en abattoirs (température, humidité relative, pratiques de nettoyage/désinfection), issues des campagnes de prélèvements et d'enquêtes auprès des professionnels,
- des expérimentations en laboratoire visant à évaluer l'impact de l'humidité relative, de la température, des milieux de croissance et de traitements lumineux (LED) sur la survie, la résistance et l'inactivation des formes végétatives et sporulées des deux pathogènes.

Les données produites sont exploitées par des analyses expérimentales, statistiques et, le cas échéant, de modélisation afin d'aboutir à des recommandations opérationnelles et des propositions de mesures de maîtrise.

Aucune donnée à caractère personnel n'est collectée dans ce WP.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Les données produites comprennent :

- données environnementales issues des abattoirs (température, humidité relative),
- résultats expérimentaux de croissance, survie, résistance et viabilité bactérienne,
- données issues d'analyses cellulaires et moléculaires.

Les formats utilisés sont principalement des **fichiers tabulaires** (XLSX, CSV), des **documents de résultats** (PDF) et des **images** issues des analyses.

Les volumes de données sont modérés et compatibles avec les capacités de stockage des partenaires.

2. Documentation et qualité des données

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

Les données sont associées à des métadonnées décrivant :

- les conditions expérimentales (température, humidité relative, supports, traitements),
- les souches utilisées,
- les dates et responsables des expérimentations.

Les protocoles expérimentaux sont documentés dans les cahiers de laboratoire et les fichiers associés afin d'assurer la traçabilité et la reproductibilité.

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

La qualité des données est assurée par :

- l'utilisation de protocoles standardisés,
- la réalisation d'expériences répétées et de témoins,
- le respect des procédures qualité et de métrologie des équipements.

3. Stockage et sauvegarde pendant le processus de recherche

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Les données sont stockées :

- sur les serveurs sécurisés des institutions partenaires,
- dans les cahiers de laboratoire et fichiers numériques associés,
- et partagées, si nécessaire, via l'espace collaboratif sécurisé RESANA.

Des sauvegardes régulières sont assurées par les services informatiques institutionnels.

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

L'accès aux données est restreint aux personnes impliquées dans le projet.

Les données produites ne sont pas considérées comme sensibles et ne contiennent pas de données à caractère personnel.

4. Exigences légales et éthiques, codes de conduite

4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Les données à caractère personnel, lorsqu'elles existent, sont anonymisées conformément à la réglementation en vigueur (RGPD).

4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Les droits de propriété intellectuelle et les modalités de gestion des données sont définis dans l'accord de consortium signé

par l'ensemble des partenaires.

Conformément à la loi n°2016-1321 pour une République numérique, les données issues d'une recherche financée sur fonds publics ont vocation à être rendues accessibles, sous réserve de l'absence de contraintes juridiques ou de confidentialité.

4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Le projet ne soulève pas de problématiques éthiques spécifiques.

5. Partage des données et conservation à long terme

5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?

L'ensemble des données sont partagées entre les partenaires du projet via l'espace sécurisé RESANA.

Les prélèvements réalisés lors des campagnes de prélèvement sont détruit après analyse.

Les données concernant le nom des établissements participants resteront confidentielles et une anonymisation des données de séquences humaines sera respectée.

Les données de valorisation (avancements du projet, affiches, conférences, articles ...) seront accessibles progressivement, au fur et à mesure de leur intégration sur le site internet tout au long du projet. Nous souhaitons rendre ces données accessibles le plus rapidement possible afin de rendre visible l'état d'avancement du projet et dans les principes de la science ouverte.

L'ensemble des données collectées non confidentielles dans le cadre de ce projet seront disponibles en libre accès après l'embargo de publication finale des données traitées comme indiqué dans l'accord de consortium.

5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

Pour le moment, les partenaires du projet s'engagent à conserver l'ensemble des données/métadonnées produites sur les serveurs de leurs instituts respectifs pendant toute la durée du projet et à minima pendant 3 années supplémentaires.

Tous les cahiers de laboratoire papier seront conservés et archivés.

Les données génomiques déposées dans Genbank seront préservées dans le cadre des missions du NCBI (<https://www.ncbi.nlm.nih.gov/genbank/>).

Plus particulièrement, pour les partenaires :

Anses Maisons-Alfort :

L'ensemble des souches de *C. perfringens* envoyées par les partenaires seront conservées à -80°C.

Les données collectées sont conservées sur le réseau de l'Anses.

Anses Ploufragan (filière volailles) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés en DASRI lorsque la mise en collection des isolats est terminée. Les isolats sont stockés au surgélateur. Pour *C. perfringens*, deux tubes sont préparés par isolat : un est stocké à Ploufragan, le second est envoyé à Maisons-Alfort. Pour *C. difficile*, les isolats sont repiqués sur ChromID à partir du tube stocké au surgélateur puis envoyé à l'hôpital Saint-Antoine (WP3).

Les prélèvements destinés à l'études des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant extraction et séquençage (WP6).

les données de séquençage seront déposées dans des bases publiques (SRA).

Les données collectées sont conservées sur le réseau de l'Anses.

Institut Pasteur :

Les données de séquençage des souches de *C. perfringens* et *C. difficile* sont stockées sur le réseau de l'Institut Pasteur.

Hôpital Saint-Antoine :

Archivage papier (feuilles de paillasse + CR) dans le bureau du CNR *Clostridioides difficile* au sein de l'Hôpital St Antoine.
Archivage informatique (réseau APHP + disque dur externe pour les fichiers .fsa + Résana + site Pasteur)

UMR PAM :

L'ensemble des données produites seront conservées sur le long terme par un cahier de laboratoire et un disque dur externe. Le cahier de laboratoire et le disque dur externe seront conservés à vie dans des armoires fermant à clé dans l'unité de recherche.

IDELE (filière bovine) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés lorsque la confirmation des souches isolées (*C. perfringens* par ANSES et *C. difficile* par Hôpital Saint-Antoine) a été réalisée. Les souches isolées sont conservées à -80°C en parallèle de l'envoi à l'ANSES Maisons-Alfort et à l'Hôpital Saint-Antoine (WP3). Les prélèvements destinés à l'étude des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant envoi à l'ANSES de Ploufragan (WP6).

Pour le WP3, les souches potentielles issues de la filière bovine sont identifiées de la manière suivante (commune avec les autres partenaires) : CBAT-P-B-XXX (ClostAbat - Perfringens - Bovin - numéro de 1 à XXX) et CBAT-D-B-XXX (ClostAbat - Difficile - Bovin - numéro de 1 à XXX).

IFIP (filière porcine) :

Les prélèvements réalisés lors des campagnes (WP2) sont éliminés après traitement lors de chaque campagne. Les souches isolées sont conservées à -80°C en parallèle de l'envoi à l'ANSES et à l'hôpital Saint-Antoine (WP3).

Les prélèvements destinés à l'étude des microbiotes (fèces, ambiance, etc.) sont conservés à -80°C avant envoi à l'ANSES Ploufragan (WP6).

Les données obtenues sont conservées dans l'espace de stockage des données de l'IFIP par archivage

5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

Les outils logiciels pour accéder et utiliser les données/métadonnées du projet sont standards et ouverts pour faciliter le partage et la réutilisation à long terme des données.

Néanmoins, certains logiciels utilisés pour les analyses sont restreints d'accès :

- GeneMapper : logiciel installé sur les ordinateurs du CNR *Clostridioides difficile*. Code de connexion (login + mot de passe) unique pour l'équipe du CNR
- Webribo pour la lecture des fichiers .fsa (données brutes) : site internet. Chaque nouvel utilisateur doit se créer son compte. Pas de visualisation possible des données des autres comptes (mais possibilité de retélécharger le fichier .fsa de la souche issue du séquenceur).
- BioNumerics : logiciel d'analyse des profils de typage moléculaire M13 pour les souches de *C. perfringens*.

5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?

Il n'est pas envisagé actuellement d'attribuer systématiquement de DOI à la production d'un jeu de données. Cette attribution reste à la

charge du responsable scientifique de chaque partenaire du projet réalisant des analyses.

Chaque donnée est identifiée par une numérotation particulière définie pour chaque partenaire en charge des prélèvements.

A savoir :

Anses Maisons-Alfort :

Dans le cadre du WP1, afin qu'ils soient faciles à retrouver et identifier, le nom du fichier de chaque rapport comporte la date et le type de réunion et dans le cadre d'une présentation les initiales de la personne qui l'a réalisé (exemple : 2023_01_15_PresentationWP1_OF).

Les données sont référencées par un code défini en interne par chaque partenaire qui permet de les identifier et les retrouver facilement. A titre d'exemple dans le WP1, le nom du fichier de chaque rapport comportera la date et le type de réunion et dans le cadre d'une présentation les initiales de la personne qui l'a réalisé (exemple : 2023_01_15_PresentationWP1_OF).

Les publications produites seront identifiées par un DOI.

Anses Ploufragan (filière volailles) :

Pour le WP2, un N° de série a été attribué au projet: 23TP01-AA (où AA est le N° de l'analyse, 23TP01-01 signifie que c'est la première mise en analyse du projet Clost'Abat). Les échantillons sont numérotés avec le code 23TP01-XX où XX est le N° de l'échantillon. Cette numérotation est celle décrite dans les documents qualité de l'unité HQPAP (PPN/PS/0200).

Hôpital Saint-Antoine :

Dans le cadre du WP4, un numéro interne de suivi de souches est attribué à chaque souche avérée *Clostridioides difficile*. Exemple : CD23-037, avec :

- CD : pour *Clostridioides difficile*
- 23 : les deux derniers chiffres de l'année en cours
- 037 : numéro d'incrémentation des souches du CNR.

Ce numéro suit la souche : tableau Excel, site Pasteur, feuilles de paillasse, fichier .fsa, compte rendu ...

UMR PAM :

Chaque nouvelle expérience est identifiée par un code selon la condition testée. Par exemple, pour les essais d'humidité relative sur les cellules végétatives, le code suivant a été utilisé : LS_CP_HR_CV001

- LS : nom de l'expérimentateur
- CP : nom de l'espèce bactérienne
- HR : humidité relative
- CV : cellules végétatives
- 001 : numéro d'expérience référencé à une date dans le cahier de laboratoire et dans un tableur

Les publications produites seront identifiées par un DOI.

IFIP (filière porcine) :

Pour le WP2, les prélèvements sont identifiés selon une nomenclature standardisée. Le n° d'échantillon est composé du code projet CBAT suivi du code Abattoir Ab, HP ou TL, du n° de campagne 1, 2 ou 3, du numéro de lot prélevé le jour de la campagne, de 1, à 4, du numéro de porc prélevé, de 1, à 5 et du type de prélèvement, F pour fèces, C pour carcasses et P pour le muscle Poitrine : CBAT-HP1-1-4-P.

pour les prélèvements d'environnements, les codes sont comme suit : Nom du projet CBAT, code abattoir, n° de campagne, A pour abattoir et D pour découpe, air 1 à air 3 pour les échantillons d'air, E1 à E5 pour les échantillons de surface: CBAT-HP1-D-E1 ou CBAT-HP1-D-air1.

Pour les souches détectées et stockées, les numéros suivent la même nomenclature que le prélèvement auquel est ajouté col1 à col5 pour colonie 1 à 5 : CBAT-Ab1-3-3-C-col4

IDELE (filière bovine) :

Pour le WP2, les prélèvements de chaque campagne sont identifiés de la même manière, avec un code spécifique pour chaque campagne. Chaque campagne est numéroté de 1 à 8 avec la localisation du prélèvement. Les fèces sont identifiés avec F, les carcasses avec C, les muscles avec M, les environnements abattoir EA, les environnement découpe ED.

Par exemple, l'échantillon de fèces de l'animal 1 de la campagne numéro 1 est noté : C1F1 : C1 (campagne 1, fèces animal numéro 1). Et ainsi de suite ...

6. Responsabilités et ressources en matière de gestion des données

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

La gestion des données (saisie, production, stockage, sauvegarde) est assurée par les responsables scientifiques de chaque partenaire, conformément aux règles de leurs institutions.

Le coordinateur du projet est responsable de la mise en œuvre et du suivi du PGD.

6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

Le financement du stockage local des données est assuré par l'institut hôte, l'Anses qui centralisera l'ensemble des métadonnées générées dans le cadre du projet.

L'accessibilité et la réutilisation des données ouvertes seront assurées par prise de contact avec le coordinateur et les responsables scientifiques de chaque WP, leurs adresses électroniques se trouvant sur le site internet dédié au projet.

Du temps dédié à la gestion des données est prévu, notamment par le coordinateur, afin de garantir :

- la qualité des données
- leur documentation
- leur conformité aux principes FAIR (Findable, Accessible, Interoperable, Reusable)

L'utilisation de formats ouverts et de standards partagés contribue à assurer l'interopérabilité et la réutilisation des données.

