
DMP du projet "Plateau technique mutualisé (PTM) "ANalyses des Acides Nucléiques" (ANAN)"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "INRA - Trame Structure" fourni par INRAE - Institut national de recherche pour l'agriculture l'alimentation et l'environnement.

Renseignements sur le plan

Titre du plan	DMP du projet "Plateau technique mutualisé (PTM) "ANalyses des Acides Nucléiques" (ANAN)"
Version	Version initiale
Domaines de recherche (selon classification de l'OCDE)	
Langue	fra
Date de création	2019-07-02
Date de dernière modification	2024-02-26
Licence	Etalab Open License 2.0

Renseignements sur le projet

Titre du projet	Plateau technique mutualisé (PTM) "ANalyses des Acides Nucléiques" (ANAN)
Acronyme	PDG_ANAN
Résumé	<p>Maison de la Recherche Campus du Végétal 42 rue Georges Morel – CS 60057 49071 Beaucouzé cedex - France</p> <p>Site internet</p> <p>Le plateau technique mutualisé "ANalyses des Acides Nucléiques" (ANAN) est l'un des trois plateaux de la SFR Quasav. Il a pour objectif de mutualiser et de mettre à disposition de la communauté scientifique de la SFR et de ses partenaires, des outils d'analyses en génomique (génomique, transcriptomique...)</p> <p>Les utilisateurs ont également à leur disposition des appareils dédiés à la préparation, au dosage et au contrôle qualité de leurs échantillons d'acides nucléiques.</p> <p>L'équipe ANAN est présente pour accompagner les utilisateurs dans le montage de leurs projets et le design de leurs expériences.</p>

Partenaires

- ANSES - Laboratoire de la santé des végétaux (201023299L)
- Groupe d'études et de contrôle des variétés et des semences (198321877H)
- Institut de Recherche en Horticulture et Semences (201220383H)
- Signalisation Fonctionnelle des Canaux Ioniques et Récepteurs (199614161P)
- SUBSTANCES D'ORIGINE NATURELLE ET ANALOGUES STRUCTURAUX (199213267K)
- Légumineuses, Ecophysiologie Végétale, Agroécologie (199522196C)
- Groupe de Recherche en Agro-alimentaire sur les Produits et les Procédés (199522194A)
- environnement physique de la plante horticole (201220788Y)
- Institut de Génétique Environnement et Protection des Plantes (201220379D)
- Unité en sciences biologiques et biotechnologies UMR 6286 (ex Unité de fonctionnalité et Ingénierie des Protéines) (201220063K)
- Biopolymères, Interactions Assemblages (199617832E)

Produits de recherche :

1. Données acquises sur instruments (Jeu de données)
2. Codes informatiques embarqués aux instruments du laboratoire (Logiciel)

Contributeurs

Nom	Affiliation	Rôles
Bahut Muriel	Université d'Angers	<ul style="list-style-type: none"> • Coordinateur du projet • Personne contact pour les données (Données acquises, Codes embarqués)
Jeauffre Julien - julien-jeauffre	INRAE	<ul style="list-style-type: none"> • Responsable du plan de gestion de données

Droits d'auteur :

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

DMP du projet "Plateau technique mutualisé (PTM) "ANalyses des Acides Nucléiques" (ANAN)"

Informations sur la structure

Nom de la structure

Plateau technique mutualisé "ANalyses des Acides Nucléiques" (ANAN)

Type de structure

- Plateforme, plateau technique

PTM ANAN
Maison de la Recherche
Campus du Végétal
42 rue Georges Morel – CS 60057
49071 Beaucouzé cedex - France

[Site internet](#)

Le plateau technique mutualisé "ANalyses des Acides Nucléiques" (ANAN) est l'un des trois plateaux de la SFR Quasav. Il a pour objectif de mutualiser et de mettre à disposition de la communauté scientifique de la SFR et de ses partenaires, des outils d'analyses en **génomique** (génotypage, transcriptomique...)

Les utilisateurs ont également à leur disposition des appareils dédiés à la **préparation, au dosage et au contrôle qualité** de leurs échantillons d'acides nucléiques.

L'équipe ANAN est présente pour accompagner les utilisateurs dans le montage de leurs projets et le design de leurs expériences.

Identifiant de la structure

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

Responsabilités dans la structure

Nom, Prénom	Courriel	Rôle
Balzerque, Sandrine	sandrine.balzerque@inrae.fr	Responsable scientifique
Bahut, Muriel	muriel.bahut@univ-angers.fr	Responsable technique
Jeauffre, Julien	julien.jeauffre@inrae.fr	Responsable Plan de gestion de Données

Etablissement(s) tutelle(s)

Université d'Angers

Département de rattachement Inra

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

- INRAE
- Université de Bordeaux
- Région Nouvelle Aquitaine
- Etat Français
- GIS-IBiSA
- CNOC
- Recettes

Informations sur le plan de gestion

DOI *(version publiée du plan de gestion)*

Pas encore publié

Historique des versions

Date	n° de version	Status	Auteur	Affiliation de l'auteur <i>(se reporter à l'annuaire Inra)</i>	Validé par	Validé le
19/02/2024	0	en cours	Julien JEAUFFRE			

Présentation générale des données

Données acquises sur instruments

Mode d'obtention des données

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

Les données expérimentales sont obtenues grâce aux instruments de mesure. Elles peuvent être brutes (sortie de l'instrument), et/ou bien transformées (transformation informatique)

- Les données obtenues grâce aux instruments de contrôle-qualité sont des données brutes.
 - Données chiffrées de dosage de molécule (technologies par absorbance & fluorimétrie)

- Données de profil électrophorétique

- Les données obtenues grâce aux instruments de séquençage haut débit sont des données brutes et raffinées.
 - Données court fragment (technologie Illumina)
 - Données long fragment (technologie Oxford Nanopore)
- Les données obtenues grâce aux instruments d'amplification (thermocycleurs) sont des données brutes.
 - Données de géotypage SNP (technologie KASP)
 - Données d'expression (par sonde et par colorant).

Lien vers schéma illustrant le PGD (sur site web pgtb quand le pgd sera publié)

Origine

- Expérimentation

Instruments de préparation des acides nucléiques (manipulation de solutions , extraction d'acides nucléiques, fragmentation, fabrication de banques NGS)

- Robot Perkin-Elmer, Zephyr
- Robot ID-Solutions, IDEAL-32
- Sonicateur Covaris, M220

Instruments de contrôle qualité

- Spectromètre Thermo-Scientific, NanoDropOne
- Lecteur de plaque BMG-Labtech, FluoStar
- Electrophorese capillaire automatisée Agilent Bioanalyser 2100

Instruments d'amplification

- Thermocycleur T100 (Bio-Rad)
- Thermocycleur en temps réel CFX Touch (Bio-Rad)
- Thermocycleur en temps réel CFX OPUS (Bio-Rad)

Instruments de séquençage haut débit:

- iSeq 100 (Illumina)
- MiSeq (Illumina)
- Minion Mk1B (Oxford Nanopore Technology)
- Minion Mk1C (Oxford Nanopore Technology)

Type de données

- Dataset
- Workflow
- Other (à préciser dans la zone "Informations supplémentaires)

Dataset : Fichiers de données expérimentales ou de traitements de données

Workflow : Séquence de traitement de données (Analyse SSR-Seq, Métabarcoding principalement)

Autres:

- Programmes de manipulation robotisée (Robot Perkins-Elmer Zephyr, Robot IDEAL32)
- Programme de mesure automatisée (Lecteur en plaque BMG-FluoSTAR)

Nature des données

Données génomiques expérimentales

- Donnée chiffrée quantifiant les solutions d'acide nucléiques
- Profil électrophorétique qualifiant les solutions d'acide nucléiques
- Donnée chiffrée post-amplification (Cq) qualifiant les solutions d'acide nucléiques
- Séquence nucléotidique (ADN/ARN)

Données dérivées ou compilées, issues de traitement et de l'analyse des données brutes

- Contrôle qualité (QC, Image, Témoins)
- Analyse de données -omique (génomique, transcriptomique, métagénomique)
- Rapport des données et des analyses
- Système d'information (Redmine, lien interne <http://147.100.175.60/>)

Format des données

Type de données	Format informatique de sortie (propriétaire / ouvert)	Format informatique transformé (propriétaire / ouvert)
Données chiffrées de quantification des acides nucléiques par mesure d'absorbance ou de fluorimétrie	Comma-separated values (csv) Tab-separated values (tsv) Structured Query Language (.sql)	
Données de qualification des acides nucléiques par profil électrophorétique	.xad	Portable Document Format (pdf)
Données de séquençage massif	lecture courte (Illumina): Binary Base Call (BCL) lecture longue (Oxford Nanopore): Hierarchical Data Format (.fast5)	fastq (fastq; fq) GNU zip (.gz)
Données d'amplification * génotypage SNP chimie KASP * expression par chimie intercalante ou sonde FRET	Data (pcrd) Gene Study (mgxd) Fichier de pré-données autonome (zpcr)	/

Multiples formats de données, dépendant de l'origine des données

- Séquençage : fasta, , md5sum, zip, html
- Génotypage : xml, xls, png
- Analyse bioinformatique : sam, bam, fasta, fastq, vcf, gff, bed, txt, csv, html, xls, Rdata
- Programmation : Python, Bash, Awk, Perl, R, Snakemake, Yaml, Git, Markdown
- Texte : pdf, odf, doc, docx, txt
- Dataset : xls, xls, txt, csv
- Rapports : pdf, doc, docx, ppt, xls, xls, html, tar.gz, png

Périmètre thématique des données

- Biodiversity and Ecology
- Food Safety and Toxicology
- Forests and Forest Products
- Microorganisms
- Omics
- Plant Breeding and Plant Products
- Plant Health and Pathology
- Soils and soil sciences

Codes informatiques embarqués aux instruments du laboratoire

Mode d'obtention des données

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

Un jeu de codes informatiques pilotant l'instrument de manipulation robotisée ID-Solutions IDEAL32, est fourni par défaut par le fournisseur de l'instrument (Macherey/Nagel).

Un jeu de codes informatiques pilotant l'instrument de mesure automatisée BMGLabtech-FluoSTAR, et des scripts d'automatisation de ces codes sont fournis par défaut par le fabriquant de l'instrument .

Ces codes de pilotage et scripts d'automatisation font l'objet de modifications pour les adapter aux spécificités des échantillons biologiques et/ou questions de recherche. Ces modifications sont tracées et versionnées.

Des codes informatiques pilotant l'instrument de manipulation robotisée Perkins-Elmer Zephyr sont conçus en interne. Ils font l'objet de modifications pour les adapter aux spécificités des échantillons biologiques et/ou questions de recherche. Ces modifications sont tracées et versionnées.

Les codes informatiques réalisant l'analyse secondaire et tertiaires des données.

- Génomique : Assemblage, appel de variants
- Transcriptomique : Analyse d'expression différentielle
- Métagénomique : Métabarcoding

Origine

- Code

Logiciels commerciaux (Epi2me, BMG-LabTech OMEGA)

Type de données

- Software

Nature des données

Codes informatiques

Format des données

Type de données	Format informatique de sortie (propriétaire / ouvert)	Format informatique transformé (propriétaire / ouvert)
Programme de mesure automatisée * Quantification des acides nucléiques * Quantification des activités enzymatiques * Quantifications des métabolites primaires et secondaires	LecteurBMG-FluoSTAR: code (.tsc) script (.btc)	Excel Binary File Format (xls) Extensions to the Office Open XML SpreadsheetML File Format(xlsx) Comma-separated values (csv)
Programme d'amplification * génotypage SNP chimie KASP * expression par chimie intercalante ou sonde FRET * Fabrication de banque destinée au séquençage massif	Protocol (prcl) Plate (.pltd)	
Programmes de manipulation robotisée * Extraction des acides nucléiques * Manipulation de solutions	Robot Perkins-Elmer Zephyr: Robot ID-solutions IDEAL32: .txt /	

Périmètre thématique des données

Question sans réponse.

Droits de propriété intellectuelle

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

- **Projet en Prestation** : la PGTB est propriétaire des méthodes et l'utilisateur est propriétaire des résultats. La PGTB peut donc librement communiquer sur les méthodes, mais ne pourra communiquer sur les résultats qu'après accord écrit de l'utilisateur. En terme de valorisation, l'utilisateur doit systématiquement mentionner la PGTB dans les remerciements de toutes les valorisations scientifiques, en cas d'une implication significative de la PGTB, le co-autorat sera de mise.
- **Projet en R&D/Collaboration** : l'utilisateur et la PGTB définissent la propriété intellectuelle des méthodes et des résultats d'un commun accord dans le cahier des charges. Il en sera de même dans le cas de valorisation scientifique.

Confidentialité

Identification des jeux de données contenant des données confidentielles

De manière générale, les données sont la propriété des utilisateurs et sont considérées comme des informations confidentielles. Les données resteront confidentielles jusqu'à publication par l'auteur.

Dans le cas spécifiques des projets internes, les données ont vocations à être rendues publiques (publications, poster, data.gouv).

La liste des clients et données à caractère personnel sont listés en interne à la structure.

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

Anonymisation des noms de projets et des échantillons proposés aux utilisateurs.
Transfert des données uniquement à l'utilisateur.
Accès aux données limité à l'utilisateur et collaborateur

Le cas échéant, comment la confidentialité de données fournies par des personnes sera garantie lorsque les données seront partagées ou rendues disponibles pour une analyse de second niveau ?

Pas concerné
Un accès aux données à un tiers sera possible si l'utilisateur porteur du projet autorise le partage de ses données pour analyse.

Partage des données

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

La donnée brute et raffinée appartient à l'utilisateur, donc le choix lui revient.
Les codes informatiques ayant concourus à l'acquisition des données seront rendus publiques au travers des forges institutionnelles.
Les données générées dans le cadre d'un projet interne financé en partie ou en totalité pour un financeur public suit la politique d'ouverture des données selon le principe "ouvert autant que possible, fermé autant que nécessaire".

Quelles sont les réutilisations potentielles de ces données ?

L'utilisateur est libre de faire ce qu'il veut des données produites (analyse, publication...)

Les données de sortie de séquençage sont des données brutes intéropérables. En ce qui concerne le traitements de données, les versions de logiciels, les machines utilisées et le cheminement des analyses sont tracés sur Redmine. Donc le travail est reproductible. De plus la majeure partie des prestations possèdent un mode opératoire décrivant le cheminement de la prestation ce qui permet de rendre réutilisable les différentes prestations effectuées

La lecture des données nécessite-t-elle le recours à un logiciel ou un outil spécifique ? Si oui, lequel ?

La lecture des données acquises grâce aux thermocycleurs ainsi que les programmes d'amplification nécessite l'usage du logiciel propriétaire CFX Maestro 2.3 (Bio-Rad)
La lecture des données aux formats Excel Binary File Format (xls) et Extensions to the Office Open XML SpreadsheetML File Format(xlsx) nécessite l'usage du logiciel propriétaire MicroSoft Excel
La lecture des données de qualification des acides nucléiques par profil électrophorétique nécessite l'usage du logiciel propriétaire 2100 Expert (Agilent)
L'usage de ces trois logiciels propriétaires oblige l'utilisateur à acheter une licence commerciale s'il souhaite analyser les données brutes sur son poste informatique

Comment les données seront-elles partagées ?

Les données acquises par les instruments de contrôle-qualité & les instruments de séquençage haut débit sont mises à disposition sur le serveur de stockage NAS du pôle informatique de l'IRHS (partage-bioinfo.angers-nantes.inrae.fr/anandata), sous condition d'accès.

Les données acquises par les instruments de contrôle-qualité pour/par utilisateurs du laboratoire IRHS sont également accessibles par le serveur de stockage NAS de l'unité (pnas1.stockage.inrae.fr/angersnantes-irhs/irhs), sous condition d'accès.

Les données acquises par les thermocycleurs sont accessibles par support physique portatif (clé USB/Disque Dur)

Cas particulier des données acquises par l'instrument de séquençage haut débit Illumina NextSeq550:

Les données acquises sont accessibles par support physique portatif (Disque Dur) puis mises à disposition sur le serveur de stockage NAS du pôle informatique de l'IRHS (partage-bioinfo.angers-nantes.inrae.fr/anandata), sous condition d'accès.

- Envoi des données par Filesender. (<https://filesender.renater.fr/>)
- Utilisation de la solution de stockage Illumina Basespace

La mise à disposition des données brutes et raffinées dans les entrepôts spécialisés en génomique auprès de la communauté scientifique nationale et internationale reste de la responsabilité de l'utilisateur selon les moyens jugés pertinents.

Avec qui ?

- Autre

Les données sont uniquement partagées avec l'utilisateur.

Sous quelle licence ?

Non concerné

Organisation et documentation des données

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

Utiliser éventuellement un lien vers un schéma illustrant les processus

- Définition des besoins et des méthodes avec l'utilisateur
- Réception des échantillons
- Validation des échantillons (respect des conditions d'acceptabilité : conditions générales, prérequis)
- Stockage des échantillons
- Préparation des échantillons (dosage, dilution, amplification ...)
- Séquençage ou génotypage des échantillons
- Stockage et archivage des données sur le NAS
- Contrôle qualité
- Analyse bio-informatique (si dans la prestation)
- Stockage et archivage des résultats/données traités sur le NAS.
- Traçabilité des analyses réalisées (cheminement) et des métadonnées pour chaque projet sur PGTBsi-Redmine (appareil utilisé, dates, échantillons, manipulateur, réactifs, suivi d'utilisation ...). Pour chaque appareil ou traitement de données, des tableaux ou documents de suivi exhaustif sont rédigés (date, opérateur, lots de réactifs, contrôles qualité, ...).
- Rédaction rapport pour les prestations de bio-informatiques en séquençage, ou envoi d'un mail.
- Envoi des données, des résultats issus du traitement de données et rapport d'analyse à l'utilisateur.
- Destruction des échantillons et des données 2 mois après l'envoi des derniers résultats (sauf projets R&D et internes)
- Possible interaction avec l'utilisateur (réunion, rédaction échange de mail) si besoin d'informations supplémentaires ou pour

valorisation scientifique.

Lien vers schéma illustrant le PGD

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

Quel que soit le projet, l'ensemble des métadonnées sont écrites et conservées sur redmine. Chaque étape de l'analyse possède des métadonnées bien spécifique à l'étape en question. De plus les machines peuvent créer leur propres fichiers de métadonnées qui peut être conservées sur le NAS ou sur Redmine.

Métadonnées	Origine, mode de production des métadonnées (ex : saisie manuelle, annotation automatique...)	Standard, Vocabulaires associés	Conditions ou fréquence de la mise à jour (si applicable) (ex : changement de l'accessibilité)
metadonnées séquençage et génotypage	Saisie via l'interface web redmine	Vocabulaire controlé pour différentes métadonnées et vocabulaire machine.	

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

Description de l'expérience, des manipulations, des protocoles (expérimentaux et bio-informatiques). Une documentation des métadonnées essentielles est rendue sous forme de rapport ou de mail à l'issue de la prestation en bio-informatique. De plus les interactions restent possible avec la plateforme une fois la prestation terminée.

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

Les fichiers de données sont conservées sur un NAS et les métadonnées ainsi que le cheminement des analyses sont conservées sur un serveur redmine.

- Sur le NAS, les prestations peuvent être rangées par une ou plusieurs de ces catégories représentées par des dossiers et sous dossiers :

- Technologies (Illumina, Oxford Nanopore Technology, Agena Bioscience ...)
- Machines (NextSeq2000, Miseq ...)
- Activités (SSR-Seq, Métabarcoding, ...)
- Années

Puis dans le sous dossier le plus approprié à la prestation un dossier portant l'acronyme du projet sera créé qui contiendra l'ensemble des fichiers associées à la prestation. Chacun de ces dossiers contient les Inputs et Outputs des différents étapes réalisées dans le cadre de la prestation.

- Sur redmine un projet est créé pour chaque prestation, ce projet porte l'acronyme du projet. Puis chaque sous tâches associées à la prestation sont associées au projet, afin que toutes les métadonnées soient reliées les unes aux autres.

D'autre part, concernant le nommage des fichiers, les machines peuvent générer leur propres fichiers suivant leurs propres conventions de nommage. En ce qui concerne les fichiers créés/générés par le personnel de la plateforme, il n'y a pas de convention de nommage clairement établie, mais le personnel veille à ce que le nom du fichier soit parlant.

Enfin, concernant les versions un tableau des suivis des versions est disponible en interne sur le NAS qui est mis à jour lorsqu'une version de machine ou de logiciel change au sein de la plateforme.

Quelle est le processus de contrôle qualité des données ?

- Contrôles pour valider la conformité et la qualité des échantillons entrants (ADN/ARN) -> précisé pour chaque analyse dans des prérequis spécifiques accessibles sur notre site web (<https://pgtb.fr/documents-et-liens/>).
- Contrôles internes qui valident la qualité du séquençage et du génotypage (témoin négatif, témoin positif). Dans le cas du séquençage des outils de QC sont aussi utilisées (fastQC, PycoQC)
- Possibilité de contrôle qualité dans certaines analyses bio-informatiques.

Ces contrôles sont compilés et analysés dans le cadre des certifications ISO 9001 et NFX 50-900 sous forme de KPI et d'indicateurs.

Stockage et sécurité des données

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

La PGTB utilise un NAS pour stocker ces données.

Lorsque le séquençage ou le génotypage est terminé un copier coller des données depuis la machine vers le NAS est fait.

Les espaces de stockage de ces NAS possèdent des sauvegardes à l'heure (snapshot) pour la journée en cours et à la journée pour le mois en cours. De plus il y a un archivage des données grâce à 2 copies à 2 endroits différents fait automatiquement mensuellement.

Pour les données issues du Nextseq 2000, en plus d'être copier coller vers un NAS, les données sont aussi transférées au cloud Illumina Basespace (flux/transfert intégré à la machine).

Lien vers schéma illustrant le PGD

Quelle est la volumétrie actuelle et prévisionnelle ?

La volumétrie actuelle est de 43.3 TB sur l'espace PGTP. La volumétrie disponible est de 19.54 TB.

La volumétrie actuelle est de 46.39 TB sur l'espace NextSeqSpace/ONT_backup. La volumétrie disponible est de 37.4 TB.

En ce qui concerne la volumétrie prévisionnelle, il n'y a pas de changements à prévoir avec les équipements présent sur la PGTB, car la plateforme n'a pas vocation à faire du stockage de données. Les données des utilisateurs sont censées être supprimées 2 mois après la fin du projet.

Un changement de volumétrie serait à prévoir si la PGTB acquerrait de nouveaux équipements nécessitant des plus gros espaces de stockage (car des plus gros volumes de données produites par RUN).

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

Repose sur les SI de l'unité (UMR 1202 BIOGECO), politique de sécurité en vigueur :

- dans le portail Bordeaux Aquitaine
- Portail des données INRAE
- [Charte des infrastructures de recherche à l'Inra](#)

Cela implique une sauvegarde du NAS avec réplication

Sécurité - Confidentialité : les données font-elles l'objet d'échange ou de partage avec de tiers acteurs et selon quelles modalités ? comment sont déterminés les droits d'accès aux données avant leur publication ?

Seuls les coordinateurs ou collaborateurs du projet et le personnel de la plateforme ont accès aux données tant qu'elles ne sont pas publiques.

Pas d'échanges avec des acteurs tiers par défaut, sauf dans le cas des données NextSeq 2000 stockées sur le cloud BaseSpace d'Illumina qui peuvent être accessibles au support technique lorsqu'un run est problématique (accès sécurisé et individuel).

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

Sécurité

- Protection contre les virus et les intrusions géré par le site de Cestas Pierroton
- Restrictions sur le droit d'accès
- Sauvegarde des données : snapshots plusieurs fois par jour (toutes les heures) conservé sur 24h. Sauvegarde journalière. Copie/Archive sur serveur secondaire une fois par mois

Archivage et conservation des données

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

Le plateau ANAN est un plateau d'acquisition de données.

Dans le cadre de prestations de service, il n'a pas vocation à conserver les données produites sur le moyen/long terme. La durée choisie pour la mise à disposition des données est de trois années (après la clôture du projet (dernières données envoyées)). Les données ne sont pas conservées au-delà, seules les métadonnées le sont. Possibilité de déroger à cette règle si accord préalable avec l'utilisateur.

Dans le cadre de projets collaboratifs et/ou de projets de Développement méthodologique, les données sont conservées jusqu'à leur publication et selon les principes FAIR.

A plus long termes toutes les données sont détruites, la plateforme n'a pas vocation à faire de l'archivage de données. La conservation des données appartient à l'utilisateur.

Sur quelle plateforme d'archivage pérenne seront archivées les données à conserver sur le long terme ? Sinon, quelles procédures seront mises en place pour la conservation à long terme ?

Le plateau ANAN est un plateau d'acquisition de données.

il n'a pas vocation à conserver les données produites sur le moyen/long terme.

Quelle est la durée de conservation des données ?

La règle générale pour les prestations est que 2 mois après la clôture du projet (dernières données envoyées), les données ne sont pas conservées, seules les métadonnées le sont.

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

Le plateau ANAN n'a pas vocation à conserver les données produites sur du long terme.

