



HAL
open science

Plan de Gestion de Données de l'entité "Plateforme GeT-TRiX"

Yannick Lippi

► To cite this version:

| Yannick Lippi. Plan de Gestion de Données de l'entité "Plateforme GeT-TRiX". INRAE. 2026. <hal-05565187>

HAL Id: hal-05565187

<https://hal.science/hal-05565187v1>

Submitted on 24 Mar 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.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

Plan de l'entité "Plateforme GeT-TRiX"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "Science Europe : modèle structuré pour les entités de recherche" fourni par DMP OPIDoR.

Renseignements sur le plan

Titre du plan	Plan de l'entité "Plateforme GeT-TRiX"				
Livrable					
Version	Version finale				
Objet/périmètre du plan	<p>Ce plan décrit les divers produits de recherche générés par la plateforme GeT-TRiX dans le cadre de ses activités de prestations d'analyses transcriptomiques.</p> <ul style="list-style-type: none">• les services de mise à disposition d'équipements auprès des utilisateurs• les services de génération et analyse de données transcriptomiques proposés par la plateforme• le développement de scripts d'analyse et d'une application web				
Domaines de recherche (selon classification de l'OCDE)	Biological sciences, Computer and information sciences				
Langue	fra				
Date de création	2025-11-14				
Date de dernière modification	2026-03-24				
Licence	<table><tr><td>Nom</td><td>Creative Commons Attribution Non Commercial 4.0 International</td></tr><tr><td>URL</td><td>http://spdx.org/licenses/CC-BY-NC-4.0.json</td></tr></table>	Nom	Creative Commons Attribution Non Commercial 4.0 International	URL	http://spdx.org/licenses/CC-BY-NC-4.0.json
Nom	Creative Commons Attribution Non Commercial 4.0 International				
URL	http://spdx.org/licenses/CC-BY-NC-4.0.json				

Renseignements sur l'entité de recherche

Nom de l'entité	Plateforme GeT-TRiX
Acronyme	GeT-TRiX
Identifiant	https://toxalim.toulouse.hub.inrae.fr/nos-services/get-trix-plateforme-de-transcriptomique
Type d'identifiant	URL

Description

La plateforme GeT-TRiX est hébergée par l'unité INRAE Toxalim et elle est un des quatre sites de la de la plateforme Génome et Transcriptome (GeT, <http://get.genotoul.fr>) de la Génopole Toulouse Midi-Pyrénées ([GenoToul](#)). La plateforme GeT est labellisée par IBiSA et labellisée Infrastructure Scientifique Collective par INRAE. Elle est également membre de l'Infrastructure de Recherche Distribuée nationale INRAE Genomics. Les informations particulières du site GeT-TRiX sont accessibles sur [cette page](#).

Les services proposés incluent : le conseil et l'expertise dans la préparation de projets, des prestations de services complètes de la production jusqu'à l'analyse de données transcriptomiques, des prestations sur demandes, la mise à disposition d'équipements et la formation (initiale et continue).

Ouverte à la communauté scientifique académique, publique et privée, et située au niveau local, national ou international, la plateforme offre des solutions d'analyses transcriptomiques sans *a priori* par technologies RNA-seq, ainsi que des solutions d'études ciblées à moyen débit (PCR quantitative en temps réel) accompagnées de mise à disposition d'appareils de contrôle qualité d'échantillons.

Date de création

2011-01-01

Tutelle(s) de l'entité

Nom	Toxalim Research Centre in Food Toxicology
Acronyme	Toxalim
Identifiant	https://ror.org/0111a5077
Type d'identifiant	ROR
Identifiant RNSR	
Nom	Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement
Acronyme	INRAE
Identifiant	https://ror.org/003vg9w96
Type d'identifiant	ROR
Identifiant RNSR	
Nom	Centre Occitanie-Toulouse
Acronyme	
Identifiant	https://ror.org/02sxjwh33
Type d'identifiant	ROR
Identifiant RNSR	

Source(s) de financement

Financier	Nom du financier	Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement
	Identifiant	https://ror.org/003vg9w96
	Type d'identifiant	ROR
Identifiant du financement		
Statut du financement		Approuvé

Produits de recherche :

1. Production et analyse de données transcriptomiques (Service)

2. Appareil d'acquisition de données (Jeu de données)
3. Scripts informatiques et logiciels (Logiciel)

Contributeurs

Nom	Affiliation	Rôles
Lippi Yannick		<ul style="list-style-type: none"> • (util_equipements, presta_transcripto) • Auteur (info_code) • Chef de projet (info_code) • Gestionnaire de données • Personne contact (presta_transcripto, util_equipements, info_code) • Responsable de l'entité • Responsable de l'entité (presta_transcripto, util_equipements) • Responsable de traitement des données (selon RGPD) • Responsable du plan • Sous-traitant (selon RGPD) (presta_transcripto)
Naylies Claire	Toxalim Research Centre in Food Toxicology - https://ror.org/0111a5077	<ul style="list-style-type: none"> • Autre (presta_transcripto)
Payros Gaëlle - https://orcid.org/0009-0008-4445-6188		<ul style="list-style-type: none"> • Gestionnaire de projet (presta_transcripto)

Nom	Affiliation	Rôles
Plateforme GeT-TRiX	Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement - https://ror.org/003vg9w96	<ul style="list-style-type: none"> • (util_equipements) • Autre (util_equipements) • Collecteur de données • Collecteur de données (util_equipements) • Gestionnaire de projet (presta_transcripto) • Personne contact (presta_transcripto, util_equipements)

Budget

Type de coût (Titre)	Montant	Etape du cycle de vie
Frais de personnel	EUR	Coûts liés à la collecte/production des données - presta_transcripto
Matériel informatique	EUR	Coûts liés à la collecte/production des données - presta_transcripto
Matériel informatique (Coûts des analyses de données)	EUR	Coûts liés au traitement/analyse des données - presta_transcripto

Production et analyse de données transcriptomiques

Description des données et collecte ou réutilisation de données existantes

Description générale du produit de recherche

Nom	Production et analyse de données transcriptomiques
Nom abrégé	presta_transcripto
Description	<p>La plateforme GeT-TRiX offre a ses utilisateurs un service d'analyse transcriptomique sans a priori par technologie RNA-Seq. Pour chaque projet d'analyse, la plateforme prend en charge diverses activités pour produire des résultats d'analyse transcriptomique à partir des échantillons d'ARN fournis par les biologistes. Ces activités sont réparties en trois grandes familles :</p> <ol style="list-style-type: none">1. Produire les données<ul style="list-style-type: none">• Des dosages de contrôle qualité des échantillons d'ARN fournis par l'équipe• Convertir les ARN en librairies RNA-Seq en utilisant les protocoles internes adaptés de kits commerciaux2. Faire sous-traiter le séquençage<ul style="list-style-type: none">• Envoyer les librairies à une plateforme partenaire (INRAE GeT-PlaGe ou INRAE PGTB) qui va produire des données brutes de RNA-Seq (fichiers FASTQ)3. Produire des résultats d'analyses bioinformatiques et biostatistiques des données de séquençage RNA-Seq<ul style="list-style-type: none">• Réaliser des traitements bioinformatiques à partir des données brutes (FASTQ) pour produire une table d'abondance de gènes• Réaliser des analyses biostatistiques à partir de la table d'abondance pour produire des résultats d'analyses d'expression différentielle
Type	Service
Workpackage	

Mots clés	
Mot-clé	ARN
Vocabulaire	Thésaurus INRAE
URL	http://opendata.inrae.fr/thesaurusINRAE/c_8686
Mot-clé	ARN messenger
Vocabulaire	Thésaurus INRAE
URL	http://opendata.inrae.fr/thesaurusINRAE/c_8724
Mot-clé	ARN non codant
Vocabulaire	Thésaurus INRAE
URL	http://opendata.inrae.fr/thesaurusINRAE/c_7797
Mot-clé	expression d'ARN
Vocabulaire	Thésaurus INRAE
URL	http://opendata.inrae.fr/thesaurusINRAE/c_6343
Mot-clé	transcriptomique
Vocabulaire	Thésaurus INRAE
URL	http://opendata.inrae.fr/thesaurusINRAE/c_17034
Mot-clé	biostatistique
Vocabulaire	Thésaurus INRAE
URL	http://opendata.inrae.fr/thesaurusINRAE/c_15169
Mot-clé	bio-informatique
Vocabulaire	Thésaurus INRAE
URL	http://opendata.inrae.fr/thesaurusINRAE/c_15351
Mots clés (texte libre)	
Langue	fra
Date de publication	
Identifiant pérenne	
Contient des données personnelles ?	Oui
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Oui
Est-ce que des données existantes seront réutilisées ?	
Justification	Ce service ne réutilise pas de données existantes. Il vise à générer de nouvelles données.
Comment seront produites/collectées les nouvelles données ?	

Nom de la méthode

Génération et analyse bioinformatique et biostatistique de données RNA-seq

Description

Les librairies RNAseq sont produites en utilisant divers protocoles commerciaux adaptés à chaque besoin:

- Illumina Stranded mRNA Prep, Ligation
- Illumina Stranded Total RNA Prep, Ligatoin
- Lexogen Quantseq mRNA-Seq FWD

Les données brutes de séquençage RNA-Seq sont générées par les plateformes partenaires (GeT-PlaGe ou PGTB) et sont transmises directement à l'équipe de recherche. Les méthodes de production des données brutes de séquençage sont basées sur des séquenceurs nouvelle génération à lectures courtes (Illumina Nextseq 2000 et Element Biosciences AVITI principalement) gérés par les plateformes partenaires qui concluent un contrat de service avec le client.

La plateforme GeT-TRiX réalise ensuite des analyses secondaires sur les données brutes comme décrit dans la section 4.

Nature des données

Données expérimentales

Equipements, plateaux techniques utilisés

Titre	GeT
Identifiant	https://cat.opidor.fr/index.php/GeT
Type d'identifiant	URL
Contact	get@genotoul.fr
Titre	PGTB
Identifiant	https://cat.opidor.fr/index.php/PGTB
Type d'identifiant	URL
Contact	pgtb@inrae.fr

Coûts

Type de coût	Frais de personnel
Devise	EUR
Titre	
Description	
Type de coût	Matériel informatique
Devise	EUR
Titre	
Description	

Documentation et qualité des données

Quelles métadonnées et quelle documentation (par exemple mode d'organisation des données) accompagneront les données ?

Description

Les données produites (données brutes, table de comptage, données normalisées et résultats d'analyses statistiques) sont accompagnées des informations qui décrivent à chaque étape les méthodes et matériels utilisés. Un rapport est restitué en fin d'analyse incluant une description de toutes les étapes mises en oeuvre pour générer les données (matériel et méthodes rédigé en anglais adapté au format demandé pour les articles scientifiques) ainsi que les références utilisées (kits, paramètres spécifiques, versions de logiciels, génome de référence et annotation, références bibliographiques).

Un tableur rassemblant toutes les métadonnées associées à la production et l'analyse des données est également fourni pour que l'équipe propriétaire des résultats puisse soumettre les jeux de données dans les entrepôts publics internationaux (GEO/NCBI ou ArrayExpress/EBI).

Quelles seront les méthodes utilisées pour assurer la qualité scientifique des données ?

Description

Les données brutes de séquençage (fichiers fastq) contiennent un score de qualité associé à chaque base lue (Phred score).

Dans le pipeline d'analyse de données RNAseq des métriques de contrôle qualité des données sont fournies à chaque étape pour renseigner sur la qualité des données produites (données brutes) et analysées (taux d'alignement des lectures sur génome de référence, localisation des alignements intra-génique ou inter-génique, ...). Des analyses statistiques exploratoires (ACP, classifications, distributions, corrélations, ...) sont également réalisées à chaque étape du traitement biostatistique (filtrage, normalisation) pour identifier les sources de variabilités présentes dans les données (biologique mais éventuellement techniques). Un rapport d'analyse détaillé rend compte des observations faites quant à la qualité des données et la présence éventuelle de variabilité expliquée par des facteurs techniques pouvant être prise en compte ou pas dans l'analyse statistique.

Exigences légales et éthiques, code de conduite

Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?

Description

Les questions en lien avec le traitement de données sensibles au titre du RGPD sont traitées spécifiquement dans deux cas de figure:

- **données des utilisateurs de la plateforme**: une déclaration de traitement des données des utilisateurs a été réalisée auprès de la DIL d'INRAE qui a fourni un récépissé d'enregistrement au registre de traitement.
- **données sensibles confiées par des biologistes et traitées dans le cadre de sous-traitance**: la plateforme a appliqué la fiche de recommandations rédigée par la DIL d'INRAE pour les cas de sous-traitance. La plateforme GeT-TRiX a mis en place de nombreuses procédures afin de s'engager dans la mise en conformité pour le traitement de données sensibles (échantillons humains, données de santé):
 - registre de sous-traitance stipulant le cadre du traitement, les personnes contact (DPO),
 - un contrat d'engagement réciproque sur les mesures déployées pour assurer la sécurité des données et des échanges.
 - les personnes habilités à accéder aux données
 - le chiffrage des fichiers de données sensibles et des échanges
 - la durée de conservation des données, la suppression des données

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 ?

Description

Les données générées dans le cadre de prestations sont sous la propriété des utilisateurs finaux des analyses. L'entité ne détient aucun droit de propriété sur les données générées dans le cas de prestation.

Les données générées dans le cadre de contrats de partenariat ou collaborations simples, font l'objet d'un accord de consortium. A défaut la propriété des données est partagée à tous les partenaires du contrat.

Les données sont diffusées et partagées aux seules personnes titulaires de leur propriété. Elles sont accessibles sur les serveurs de fichier du personnel de la plateforme en utilisant un contrôle d'accès de comptes.

Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?**Description**

Les aspects éthiques liés à la collecte des données en amont de l'implication de la plateforme sont de la responsabilité des demandeurs.

Un code de bonne conduite déontologique est mentionné dans l'offre de prestation afin de préciser la contribution des agents de la plateforme dans les produits de valorisations scientifiques (remerciements, co-autorat, mention dans les matériels et méthodes, ...) lorsque cela est requis.

Traitement et analyse des données**Comment et avec quels moyens seront traitées les données ?**

Description

Les données brutes (FASTQ) produites par les plateformes partenaires sont transmises à la plateforme GeT-TRiX pour la réalisation des analyses de données.

Dans ce cadre les données brutes sont analysées en utilisant les ressources de calcul de la plateforme Genotul Bioinfo hébergées au mésocentre de calcul CALMIP.

L'analyse des données comprend:

1. **analyse bioinformatique** en utilisant des pipelines publiés sur l'entrepôt public <https://nf-co.re/rnaseq>. Ces pipelines sont paramétrés selon l'étude (organisme de référence, protocole de séquençage).
2. **analyse biostatistique** en utilisant le langage de programmation R et des packages adaptés aux analyses transcriptomiques (limma, edgeR, ...) permettant le filtrage des données, l'estimation des facteurs de normalisation et l'analyse différentielle pour la recherche de gène différentiellement régulés. Des analyses multivariées (double classification, ACP, PLS-DA) sont également appliquées selon les cas de figure. Enfin des analyses d'enrichissement fonctionnel ou GSEA sont également réalisables si les annotations de gènes sont accessibles pour l'organisme étudié.

Equipements, plateaux techniques

Titre	GenoToul Bioinfo
Identifiant	https://cat.opidor.fr/index.php/GenoToul Bioinfo
Type d'identifiant	URL
Contact	https://bioinfo.genotoul.fr/index.php/about-us/contact-us/
Titre	CALMIP
Identifiant	https://cat.opidor.fr/index.php/CALMIP
Type d'identifiant	URL
Contact	admin.calmip@univ-toulouse.fr

Coûts

Type de coût	Matériel informatique
Devise	EUR
Titre	Coûts des analyses de données
Description	<p>Le coût des analyses de données sont définis en tenant compte plusieurs éléments:</p> <ul style="list-style-type: none">• Le coût des contrats de location des espaces disques auprès de la plateforme partenaire Bioinfo Genotoul• Le coût du contrat d'hébergement d'une machine virtuelle utilisée pour héberger l'application web matrix.• Les coûts d'utilisation du matériel informatique des agents impliqués dans les analyses de données• Les coûts de formation du personnel et de maintenance de l'infrastructure informatique

Stockage et sauvegarde des données pendant le processus de recherche

Comment les données seront-elles stockées et sauvegardées tout au long du projet ?

Besoins de stockage

Les données brutes de séquençage sont gérées par les plateformes partenaires qui les produisent sous leur responsabilité. L'utilisateur final demandeur de l'analyse engage ensuite sa responsabilité concernant la gestion des données une fois qu'elles lui sont transférées.

Durant les analyses de données, ces dernières (données brutes, intermédiaires et finales) sont stockées sur un espace du serveur Genobioinfo (plateforme Bioinfo Genotoul) dédié loué par la plateforme GeT-TRiX. Les données présentes sur cet espace ne présentent pas de risque en cas de perte car les données brutes initiales pourront être récupérées auprès de la plateforme qui les a produites ou bien auprès de l'utilisateur final à qui elles ont été transmises. Les analyses pourront alors être réalisées à partir de ces données brutes.

Dans le cas d'analyses en partenariat de recherche (projets portés par la plateforme ou en collaboration avec une équipe de Toxalim), un autre espace de stockage sauvegardé et répliqué physiquement est utilisé (espace loué auprès de la plateforme Bioinfo Genotoul). En cas d'incident la plateforme Bioinfo Genotoul peut être contactée pour récupérer les sauvegardes.

Une fois l'analyse des données terminée sur les serveurs de la plateforme Bioinfo Genotoul, les données générées (table de comptage, données normalisées, résultats d'analyse et graphiques) sont copiés sur l'espace dédié aux agents de la plateforme GeT-TRiX (sous active directory) sur le serveur pnas2.stockage.inrae.fr sous accès sécurisé et sauvegardé (fenêtre glissante de 4 semaines) géré par le data center INRAE Toulouse, répliqué sur le data center INRAE Ile de France. La plateforme s'engage dans ses conditions générales associées à l'offre de tarif, à conserver les données durant une année à compter de la restitution des résultats.

La volumétrie associée est d'environ 1 à 2 To par an. Les données brutes (les plus volumineuses) sont supprimées un an après le rendu des résultats pour préserver les espaces de stockage.

Volume estimé des données

2

Unité

To

Mesures prises pour la sécurité des données

Les données sont accessibles uniquement aux agents de la plateforme GeT-TRiX, que ce soit sur le serveur de de la plateforme Bioinfo Genotoul (gestion des droits des comptes utilisateurs) ou bien sur le serveur du data center INRAE Toulouse (gestion des accès via active directory, LDAP).

Partage des données et conservation à long terme

Comment les données seront-elles partagées ?

Modalités de partage

Les données analysées par la plateforme GeT-TRiX sont transmises à l'utilisateur à la fin de l'analyse avec l'aide du service FileSender (<https://filesender.renater.fr/>). Dans le cadre de prestations de services, les données générées étant la propriété de l'utilisateur, les conditions de partage et publication des données sont définies par ce dernier et non par la plateforme GeT-TRiX.

La plateforme fournit malgré tout à l'utilisateur final (demandeur de prestation ou partenaire de collaboration) l'ensemble des informations en sa possession permettant de publier les jeux de données générés dans les entrepôts publics dédiés aux données transcriptomiques (GEO/NCBI ou ArrayExpress/EBI).

Dans le cadre de collaborations de recherches la plateforme peut réaliser la soumission des données sur les entrepôts publics en commun accord avec les partenaires du projet. La plateforme publie les données générées dans le cadre de projets qu'elle porte, sous forme de jeux de données dans des entrepôts, de data paper ou d'articles scientifiques.

Potentiel de réutilisation

Comment les données seront-elles conservées à long terme ?

Justification

La plateforme s'engage dans ses conditions générales associées à l'offre de tarif, à conserver les données durant une année à compter de la restitution des résultats. Aucun système d'archivage n'est mis en place dans le cadre de prestations.

Les données produites pour le compte d'équipes sont conservées par les équipes.

Les données produites dans le cadre de projets où la plateforme est porteuse du contrat sont conservées sans limite de temps sur les espaces de stockage sus-mentionnés.

Le recours à une solution d'archivage S3 (stockage froid) est en cours de déploiement pour stocker les données des projets clôturés portés par la plateforme.

Date de début**Date de fin****Dispositions finales**

Appareil d'acquisition de données

Description des données et collecte ou réutilisation de données existantes

Description générale du produit de recherche

Nom	Appareil d'acquisition de données
Nom abrégé	util_equipements
Description	<p>La plateforme GeT-TRiX offre un service de mise à disposition d'équipements d'acquisition de données ouvert à ses utilisateurs après formation. Des données sont produites spécifiquement selon chaque type d'équipement:</p> <ul style="list-style-type: none">• Dosage et contrôle de l'intégrité des échantillons d'ARN : TapeStation 4200 et RNA/DNA ScreenTapes (Agilent Technologies, CA, USA)• Quantification des échantillons d'ARN : DropSense 96 et l'utilisation de Lunatic Plates S (Unchained Labs, CA, USA)• Analyses PCR quantitative en temps réel : QuantStudio 5 (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA)• Séquençage NGS bas débit : iSeq 100 (Illumina, CA, USA)
Type	Jeu de données
Workpackage	

Mots clés	Mot-clé	ADN
	Vocabulaire	Thésaurus INRAE
	URL	http://opendata.inrae.fr/thesaurusINRAE/c_9061
	Mot-clé	ARN
	Vocabulaire	Thésaurus INRAE
	URL	http://opendata.inrae.fr/thesaurusINRAE/c_8686
	Mot-clé	PCR quantitative
	Vocabulaire	Thésaurus INRAE
	URL	http://opendata.inrae.fr/thesaurusINRAE/c_16648
	Mot-clé	spectrophotométrie
	Vocabulaire	Thésaurus INRAE
	URL	http://opendata.inrae.fr/thesaurusINRAE/c_16140
	Mot-clé	quantification
	Vocabulaire	Thésaurus INRAE
	URL	http://opendata.inrae.fr/thesaurusINRAE/c_1826
	Mot-clé	méthode de dosage
	Vocabulaire	Thésaurus INRAE
	URL	http://opendata.inrae.fr/thesaurusINRAE/c_18056
	Mot-clé	donnée de séquençage
	Vocabulaire	Thésaurus INRAE
	URL	http://opendata.inrae.fr/thesaurusINRAE/c_918b0e29
Mots clés (texte libre)		
Langue	fra	
Date de publication		
Identifiant pérenne		
Contient des données personnelles ?	Non	
Contient des données sensibles ?	Non	
Est-ce que des données existantes seront réutilisées ?		
Justification	Ce service ne réutilise pas de données existantes. Il vise à générer de nouvelles données.	
Comment seront produites/collectées les nouvelles données ?		

Nom de la méthode

Description

Les données sont collectées spécifiquement selon plusieurs types d'applications technologiques:

- Dosage et contrôle de l'intégrité des échantillons d'ARN : TapeStation 4200 et RNA/DNA ScreenTapes (Agilent Technologies, CA, USA)
- Quantification des échantillons d'ARN : DropSense 96 et l'utilisation de Lunatic Plates S (Unchained Labs, CA, USA)
- Analyses PCR quantitative en temps réel : QuantStudio 5 (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA)
- Séquençage NGS bas débit : iSeq 100 (Illumina, CA, USA)

Nature des données

Données expérimentales

Equipements, plateaux techniques utilisés

Titre	GeT
Identifiant	https://cat.opidor.fr/index.php/GeT
Type d'identifiant	URL
Contact	get@genotoul.fr

Documentation et qualité des données

Quelles métadonnées et quelle documentation (par exemple mode d'organisation des données) accompagneront les données ?

Description

- Contrôle de l'intégrité des échantillons d'ARN ou d'ADN : version de logiciel, numéro de série de l'équipement, date de l'analyse, numéro de lot des cassettes de réactifs (screentapes), informations fournies par l'utilisateur (identification des échantillons)
- Quantification des échantillons d'ARN : version de logiciel, numéro de série de l'équipement, date de l'analyse, numéro de lot des consommables (Lunatic plates), informations fournies par l'utilisateur (identification des échantillons)
- Analyses PCR quantitative en temps réel : version de logiciel, numéro de série de l'équipement, date de l'analyse, méthodes d'acquisition utilisées, informations fournies par l'utilisateur (identification des échantillons, gènes cibles)
- Séquençage NGS bas débit : version de logiciel, numéro de série de l'équipement, date de l'analyse, numéro de lot des cassettes de réactifs et flowcell, informations fournies par l'utilisateur (identification des échantillons)

Quelles seront les méthodes utilisées pour assurer la qualité scientifique des données ?

Description

Des procédures de maintenances internes et/ou externes sont mises en place sur chacun des appareils de mesure. Des vérifications périodiques sont planifiées pour assurer l'entretien régulier des appareils et réaliser des tests de conformité à l'aide de consommables dédiés et des procédures d'auto-diagnostic du système.

- Contrôle de l'intégrité des échantillons d'ARN ou d'ADN : Contrat de maintenance préventive annuelle. Standards internes utilisés à chaque analyse
- Quantification des échantillons d'ARN : Contrôle et test de conformité annuel interne
- Analyses PCR quantitative en temps réel : Contrat de maintenance préventive annuelle. Diagnostic et calibration réalisables en interne
- Séquençage NGS bas débit : Entretien périodique interne et procédure d'auto-diagnostic planifiée

Exigences légales et éthiques, code de conduite

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 ?

Description

Les données générées sont propriété des utilisateurs finaux des appareils de mesure. L'entité ne détient aucun droit de propriété sur les données générées par ses appareils, hormis celles générées dans le cadre des projets qu'elle porte.

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

Description

Ces aspects concernent le propriétaire des données et non l'entité qui met à disposition les équipements d'acquisition.

Traitement et analyse des données

Comment et avec quels moyens seront traitées les données ?

Description

Une fois les données brutes acquises par l'appareil de mesure, des logiciels de traitement spécifiques sont utilisés pour traiter les données et aider à l'interprétation:

- Contrôle de l'intégrité des échantillons d'ARN ou d'ADN : TapeStation Software permet de piloter l'appareil et renseigner les informations sur les échantillons, puis exporter les résultats sous forme de rapports au format PDF ou de tableurs.
- Quantification des échantillons d'ARN : DropQuant pour le pilotage de l'appareil et le renseignement des informations des échantillons et des paramètres d'acquisition. cDrop pour l'analyse des données de spectre et l'estimation des abondances des composés dosés. Les résultats sont exportés dans des tableurs Microsoft Excel ou Libreoffice calc.
- Analyses PCR quantitative en temps réel : QuantStudio™ Design & Analysis Software v1.5.3 pour le pilotage de l'appareil, le renseignement des informations des échantillons et des paramètres d'acquisition.
- Séquençage NGS bas débit : iSeq Control Software, pour renseigner les informations sur les échantillons et définir les paramètres de séquençage et d'analyse pour générer les fichiers FASTQ de sortie après démultiplexage.

Stockage et sauvegarde des données pendant le processus de recherche

Comment les données seront-elles stockées et sauvegardées tout au long du projet ?

Besoins de stockage

Lors du processus d'analyse, les données sont collectées et stockées en premier lieu sur le poste informatique de pilotage de l'appareil. Sur chaque poste de pilotage un programme permet de copier automatiquement (à l'aide d'un logiciel de synchronisation) les données nouvellement enregistrées sur un espace de notre serveur de fichiers hébergé sur le datacenter du centre INRAE Toulouse. Ce serveur de fichiers est répliqué sur le datacenter INRAE Ile de France. La plateforme s'engage dans ses conditions générales associées à l'offre de tarif, à conserver les données durant une année à compter de la restitution des résultats. En plus de la duplication le serveur est sauvegardé sur une fenêtre de 4 semaines. Ainsi en cas de suppression ou corruption d'un fichier, une version antérieure peut être récupérée. La volumétrie de données dépend du type d'équipement:

- Contrôle de l'intégrité des échantillons d'ARN ou d'ADN : < 500Mo par an
- Quantification des échantillons d'ARN : < 500Mo par an
- Analyses PCR quantitative en temps réel : < 500Mo par an
- Séquençage NGS bas débit : < 5 Go / an

Volume estimé des données

7

Unité

Go

Mesures prises pour la sécurité des données

Les données sont accessibles uniquement aux agents de la plateforme GeT-TRiX, sur le serveur du data center INRAE Toulouse (gestion des accès via active directory, LDAP).

Partage des données et conservation à long terme

Comment les données seront-elles partagées ?

Modalités de partage

Les données brutes générées sur les appareils de mesure de la plateforme GeT-TRiX sont transmises à l'utilisateur automatiquement à la fin de l'analyse avec l'aide du logiciel e-SIToul DataTracker. Les données générées étant la propriété de l'utilisateur, les conditions de partage des données sont définies par ce dernier et non par la plateforme GeT-TRiX qui met à disposition seulement les équipements.

Potentiel de réutilisation

Comment les données seront-elles conservées à long terme ?

Justification

Les données brutes générées sur les appareils de mesure de la plateforme GeT-TRiX sont conservées sur une durée d'un an minimum sur le disque dur de l'appareil de mesure et sur un espace de stockage (datacenter du centre INRAE Toulouse) utilisé comme copie de sauvegarde. Les données de plus d'un an sont supprimées périodiquement pour libérer de l'espace de stockage.

Aucun archivage n'est envisagé pour ce type de données.

Volume estimé des données

7

Unité

Go

Date de début

Date de fin

Dispositions finales



Scripts informatiques et logiciels

Description générale du logiciel

Dénomination et description du logiciel

Nom complet du logiciel	Scripts informatiques et logiciels
Nom abrégé	info_code
Description	<p>Ce produit de recherche concerne deux types de produits:</p> <ul style="list-style-type: none">• Les scripts de codes utilisés pour réaliser les analyses de données. Ces codes sont utilisés par les personnels de la plateforme GeT-TRiX pour réaliser les analyses bioinformatiques et biostatistiques de données transcriptomiques dans le cadre de projets confiés à la plateforme.• Une application web RShiny nommée MATRIXapp qui est mise à disposition des biologistes pour explorer les résultats d'analyses transcriptomiques fournis.
Type	Logiciel
URL du site de présentation du logiciel (ou du projet)	https://matrix.toulouse.inrae.fr/
Date de début du développement	
Date de première version	
Principaux domaines d'application	Computer and information sciences, Mathematics, Biological sciences
Autres domaines d'application	
Numéro de version du logiciel	
Typologies du logiciel	Plug-in ou module (sous-ensemble destiné à être intégré dans un framework, dans un environnement de développement, dans un écosystème applicatif), Logiciel autonome, logiciel applicatif, outil de simulation, Outil de développement

Outils et environnement d'exécution

Environnement de développement

URL de la forge logicielle https://forge.inrae.fr/GET_TRiX/matrixanalysis; https://forge.inrae.fr/ylippi/MA_Trix_App

Mode d'accès Restreint

URL du framework d'intégration continue, s'il existe ?

Informations complémentaires

Les codes sont déposés sur la forge institutionnelle d'INRAE qui permet un suivi des versions. Un dépôt local est utilisé pour réaliser les corrections, modifications, améliorations et suivi de version de code à l'aide de l'utilitaire GIT. Une fois les modifications validées, elles sont poussées sur le dépôt distant sur forge.inrae permettant un travail collaboratif.

Documentation

Informations complémentaires

Des instructions ont été rédigées à l'attention des opérateurs analystes de la plateforme afin d'indiquer comment utiliser les codes (environnement d'exécution, paramètres, conseils et conduite à tenir).

L'usage de l'application est décrit sur l'application elle-même:
<https://matrix.toulouse.inrae.fr/>

Environnement d'exécution

Informations complémentaires

Les codes sont de deux types pour les scripts d'analyse:

- scripts Unix à exécuter sur système Linux pour la réalisation des analyses bioinformatiques
- Scripts R à exécuter dans le logiciel R V4.X sous système d'exploitation Linux ou Windows.

Les codes de l'application matrix sont R et JavaScript.

Préservation du logiciel

Référencement

Informations sur le référencement du logiciel

Identifiant pérenne (DOI, HAL Id...) <https://hal.science/hal-04251787>

Type d'identifiant HAL Id

Identifiant pérenne (DOI, HAL Id...) <https://dx.doi.org/10.17180/dj0g-5f33>

Type d'identifiant DOI

Informations complémentaires

Référencement de l'application MATRIX sur HAL et DOI attribué.

Archivage pérenne

Identifiant SWHID swh:1:dir:0da7fee03e99d06a37abe49bee1985a36413d939
Type d'identifiant SWHID

Questions juridiques

Quels sont les aspects juridiques en lien avec la gestion du logiciel ?

Licences

Nom	Creative Commons Attribution Non Commercial 4.0 International
URL	https://spdx.org/licenses/CC-BY-NC-4.0.json
Nom	GNU General Public License v3.0 or later
URL	https://spdx.org/licenses/GPL-3.0-or-later.json

Informations complémentaires

Les codes utilisés sont utilisables sous licence libre utilisation .

Valorisation scientifique

Valorisation scientifique

Question sans réponse.