
DMP du projet "Plateforme de Protéomique Clinique"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "INRAE - 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 "Plateforme de Protéomique Clinique"
Version	Version initiale
Domaines de recherche (selon classification de l'OCDE)	Biological sciences (Natural sciences)
Langue	fra
Date de création	2021-11-15
Date de dernière modification	2022-11-15
Identifiant	https://doi.org/10.57745/DUWRBR
Type d'identifiant	DOI
Licence	Creative Commons Attribution 4.0 International
Documents (publications, rapports, brevets, plan expérimental....), sites web associés	<ul style="list-style-type: none">• PPC website : https://ppc-montpellier.com/index.php/accueil-en/

Renseignements sur le projet

Titre du projet	Plateforme de Protéomique Clinique
Résumé	Analyse protéomique sur de multiple plateformes technologiques
Sources de financement	<ul style="list-style-type: none">• chu montpellier :

Produits de recherche :

1. identification de protéines en HRMS (Jeu de données)
2. dosage de molécules (RNA, métabolites) par des méthodologies basées sur de la LC-MS (Jeu de données)
3. Dosage de protéines par MRM (Jeu de données)
4. Dosage ELISA (MSD, Simoa, Ella) de molécules (Jeu de données)

Contributeurs

Nom	Affiliation	Rôles
jerome vialaret - https://orcid.org/0000-0002-3730-2366		<ul style="list-style-type: none"> • Coordinateur du projet • Personne contact pour les données (DosageMRMmolécules, Protein ID, Dosage Elisa, DosageMRMpeptide) • 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 "Plateforme de Protéomique Clinique"

Informations sur la structure

Nom de la structure

Plateforme de protéomique Clinique
CHU de Montpellier

Type de structure

- Plateforme, plateau technique

Plateforme de Protéomique Clinique
Institute for Regenerative Medicine & Biotherapy (IRMB)
Hôpital Saint Eloi
80 rue Augustin Fliche
34295 MONTPELLIER – Cedex 5
FRANCE

<https://ppc-montpellier.com/index.php/contacts/>

Identifiant de la structure

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

Question sans réponse.

Responsabilités dans la structure

Nom, Prénom	Courriel	Rôle
Hirtz, Christophe	c-hirtz@chu-montpellier.fr	Responsable scientifique
VIALARET, Jerome	j-vialaret@chu-montpellier.fr	Responsable technique
VIALARET, Jerome	j-vialaret@chu-montpellier.fr	Responsable de la conservation à long terme

Etablissement(s) tutelle(s)

CHU Montpellier

Département de rattachement INRAE (ou anciens départements Inra)

Question sans réponse.

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

CHU Montpellier
Université de Montpellier
Région Occitanie

Informations sur le plan de gestion

DOI (*version publiée du plan de gestion*)

<https://doi.org/10.57745/DUWRBR>

Historique des versions

Question sans réponse.

Présentation générale des données

identification de protéines en HRMS

Mode d'obtention des données

- Données générées par la structure

La structure va réaliser des expériences d'identification de molécules sur différentes plateformes technologiques (LC-MS et ELISA). En entrée nous avons des échantillons fournis par un collaborateur interne ou externe à la structure, qui seront préparés selon une procédure, puis analysés. La procédure de préparation de l'échantillon dépend de l'échantillon. Les informations de préparation seront reportées dans les métadatas associées aux datas. La donnée principale, est la donnée de sortie des analyseurs. C'est cette donnée qui doit être conservée, tracée.

Une convention de nommage de l'ensemble des fichiers générés au cours du projet a été adoptée.

Utilisation de règles de nommage standardisées:

-Structuration du nom de fichier : sujet, date, version

-Pas de mots vides, abréviations communément compréhensibles, dans l'idéal, ne pas répéter de termes entre noms du fichier et noms du dossier (même si dans les faits, ce n'est pas si simple...) ; **pas d'espaces** ; **pas de caractères spéciaux** (accentués, symboles... : le nom du fichier doit être lisible par des machines)

-Majuscules, underscores, caractères alphanumériques ; format de la date selon la norme ISO AAAAMMJJ

-Nombres toujours sur au moins 2 unités (en fonction du nombre de fichiers concernés: 01, 002, 0003 etc.). Le même nombre en tout cas

-Élément le plus important en premier pour faciliter la recherche du document

-Terminer par mention de version du type VP/Vdef, V0, 1, 2

-NomProjet_ObjetDuDocument_Date ou Version

MonProjet_CTreunion_20130211.pdf

Les données collectées sont sauvegardées sur le NAS inserm de notre institut. Le NAS IRMB est en Raid 5 avec Snapshot. Le protocole utilisé est CIFS avec une authentification sur un AD. L'espace de stockage sur le NAS Inserm est nominatif et visible par le groupe "PPC" définis par les administrateurs (Mr Fabrice Galia en local ou la DSI Inserm). Des accès limités et personnalisés peuvent être configurés par l'administrateur. L'ajout ou la suppression d'accès se fait sur demande à l'administrateur.

Les données sont issues des analyseurs avec le format propriétaire et transférer une fois par jour sur le NAS inserm, dans notre espace alloué.

format ".baf" généré par le logiciel de pilotage des analyseurs de la société Bruker Daltonics. Avant archivage, les données pourront être convertis en format ouvert. L'outil "MSconvert" pourra être utilisé. Il supporte la conversion des formats de sortie des analyseurs de masse vers le format standard mzML. mzML est un format ouvert qui peut être utiliser pour la soumission sur des entrepôts de données.

Origine

- Expérimentation

Différents échantillons biologiques provenant de notre Centre de Ressources Biologiques (plasma, serum, salive, CSF, urine), ou produits par nos collaborateurs (culture de cellules, animaux). Ce sont donc des nouvelles données car ces échantillons n'ont pas encore été analysés.

Les protéines sont extraites de l'échantillon. Les protéines extraites sont ensuite dénaturées, réduites, et alkylées avant une digestion enzymatique (ex: trypsine). Les peptides ainsi générés sont analysés par nanoLC-MS/MS (Impact II, Bruker Daltonics). Ces données représentent la "donnée brute". Les métadonnées associées seront documentées.

Les métadonnées seront générées/reportées par l'expérimentateur qui réalise les analyses. C'est cette personne qui sera responsable de ce jeu de métadonnées. Les métadonnées regroupent les informations sur la génération de l'échantillon (informations fournis par le collaborateur), le protocole de préparation de l'échantillon, la partie chromatographie et MS de l'analyse, et les retraitements effectués sur les données brutes.

Les vocabulaires nécessaires à chaque partie seront mis en œuvre. Sur ce site internet ([Ontology Lookup Service < EMBL-EBI](#)) sont répertoriés les vocabulaires contrôlés. Si possible, dans cette liste utilisez les vocabulaires contrôlés National Cancer Institute Thesaurus (NCIT); Experimental Factor Ontology (EFO); PRIDE PRoteomics IDentifications (PRIDE).

Type de données

- Dataset

Les données brutes sont issues d'expériences. Elles sont générées par les analyseurs. On appellera cela "la donnée brutes". Elle permet d'identifier les molécules provenant des échantillons. Ce sont des données numériques fixes.

Nature des données

Analyses LC-MS: données chromatographiques et de spectrométrie de masse.

La volumétrie est < 3Go/ analyse

Format des données

format ".baf" généré par le logiciel de pilotage des analyseurs de la société Bruker Daltonics. Avant archivage, les données pourront être convertis en format ouvert. L'outil "MSconvert" pourra être utilisé. Il supporte la conversion des formats de sortie des analyseurs de masse vers le format standard mzML. mzML est un format ouvert qui peut être utilisé pour la soumission sur des entrepôts de données.

Périmètre thématique des données

- Plant Health and Pathology
- Human Nutrition and food security
- Animal Health and Pathology
- Human Health and Pathology

dosage de molécules (RNA, métabolites) par des méthodologies basées sur de la LC-MS

Mode d'obtention des données

- Données générées par la structure

La structure va réaliser des expériences d'identification de molécules sur différentes plateformes technologiques (LC-MS et ELISA). En entrée nous avons des échantillons fournis par un collaborateur interne ou externe à la structure, qui seront préparés selon une procédure, puis analysés. La procédure de préparation de l'échantillon dépend de l'échantillon. Les informations de préparation seront reportées dans les métadatas associées aux datas. La donnée principale, est la donnée de sortie des analyseurs. C'est cette donnée qui doit être conservée, tracée.

Une convention de nommage de l'ensemble des fichiers générés au cours du projet a été adoptée.

Utilisation de règles de nommage standardisées:

-Structuration du nom de fichier : sujet, date, version

-Pas de mots vides, abréviations communément compréhensibles, dans l'idéal, ne pas répéter de termes entre noms du fichier et noms du dossier (même si dans les faits, ce n'est pas si simple...) ; **pas d'espaces ; pas de caractères spéciaux** (accentués, symboles... : le nom du fichier doit être lisible par des machines)

-Majuscules, underscores, caractères alphanumériques ; format de la date selon la norme ISO AAAAMMJJ

-Nombres toujours sur au moins 2 unités (en fonction du nombre de fichiers concernés: 01, 002, 0003 etc.). Le même nombre en tout cas

-Élément le plus important en premier pour faciliter la recherche du document

-Terminer par mention de version du type VP/Vdef, V0, 1, 2

-NomProjet_ObjetDuDocument_Date ou Version

MonProjet_CTreunion_20130211.pdf

Les données collectées sont sauvegardées sur le NAS inserm de notre institut. Le NAS IRMB est en Raid 5 avec Snapshot. Le protocole utilisé est CIFS avec une authentification sur un AD. L'espace de stockage sur le NAS Inserm est nominatif et visible par le groupe "PPC" définis par les administrateurs(Mr Fabrice Galia en local ou la DSI Inserm). Des accès limités et personnalisés peuvent être configurés par l'administrateur. L'ajout ou la suppression d'accès se fait sur demande à l'administrateur.

Les données sont issues des analyseurs avec le format propriétaire et transférer une fois par jour sur le NAS inserm, dans notre espace alloué. Avant archivage, les données pourront être converties en format ouvert.

format ".lcd" généré par le logiciel de pilotage des analyseurs de la société Shimadzu. L'outil "MSconvert" pourra être utilisé. Il supporte la conversion des formats de sortie des analyseurs de masse vers le format standard mzML. mzML est un format ouvert qui pourra être utilisé pour la soumission sur des entrepôts de données.

Origine

- Expérimentation

Différents échantillons biologiques provenant de notre Centre de Ressources Biologiques (plasma, serum, salive, CSF, urine), ou produits par nos collaborateurs (culture de cellules, animaux). Ce sont donc des nouvelles données car ces échantillons n'ont pas encore été analysés.

Les molécules sont extraites de l'échantillon brut selon différents protocoles (protocoles dépend de l'échantillon et des molécules à analyser). Une étape supplémentaire de nettoyage est souvent requise. Les molécules extraites sont analysées par LC-MRM (TQ 8060NX, Shimadzu). Ces données représentent la "donnée brute". Les métadonnées associées seront documentées pour décrire toutes les étapes de ce processus.

Les métadonnées seront générées/reportées par l'expérimentateur qui réalise les analyses. C'est cette personne qui sera responsable de ce jeu de métadonnées. Les métadonnées regroupent les informations sur la génération de l'échantillon (informations fournis par le collaborateur), le protocole de préparation de l'échantillon, la partie chromatographie et MS de l'analyse, et les retraitements effectués sur les données brutes.

Les vocabulaires nécessaires à chaque partie seront mis en œuvre. Sur ce site internet ([Ontology Lookup Service < EMBL-EBI](#)) sont répertoriés les vocabulaires contrôlés. Si possible, dans cette liste utilisez les vocabulaires contrôlés National Cancer Institute Thesaurus (NCIT); Experimental Factor Ontology (EFO); PRIDE PRoteomics IDentifications (PRIDE).

Type de données

- Dataset

Les données brutes sont issues d'expériences et sont générées par les analyseurs sont "la donnée brutes". Elle permet d'identifier/de quantifier les molécules provenant des échantillons. Ce sont des données numériques fixes.

Nature des données

Analyses LC-MS: données chromatographiques et de spectrométrie de masse.

La volumétrie est < 2Mo/ analyse

Format des données

Format ".lcd" généré par le logiciel de pilotage des analyseurs de la société Shimadzu. Avant archivage, les données pourront être converties en format ouvert. L'outil "MSconvert" pourra être utilisé. Il supporte la conversion des formats de sortie des analyseurs de masse vers le format standard mzML. mzML est un format ouvert qui pourra être utilisé pour la soumission sur des entrepôts de données.

Périmètre thématique des données

- Plant Health and Pathology
- Omics
- Microorganisms
- Human Nutrition and food security
- Human Health and Pathology
- Biodiversity and Ecology
- Animal Health and Pathology

Dosage de protéines par MRM

Mode d'obtention des données

- Données générées par la structure

La structure va réaliser des expériences d'identification de molécules sur différentes plateformes technologiques (LC-MS et ELISA). En entrée nous avons des échantillons fournis par un collaborateur interne ou externe à la structure, qui seront préparés selon une procédure, puis analysés. La procédure de préparation de l'échantillon dépend de l'échantillon. Les informations de préparation seront reportées dans les métadonnées associées aux données. La donnée principale, est la donnée de sortie des analyseurs. C'est cette donnée qui doit être conservée, tracée.

Une convention de nommage de l'ensemble des fichiers générés au cours du projet a été adoptée.

Utilisation de règles de nommage standardisées:

-Structuration du nom de fichier : sujet, date, version

-Pas de mots vides, abréviations communément compréhensibles, dans l'idéal, ne pas répéter de termes entre noms du fichier et noms du dossier (même si dans les faits, ce n'est pas si simple...) ; **pas d'espaces ; pas de caractères spéciaux** (accentués, symboles... : le nom du fichier doit être lisible par des machines)

-Majuscules, underscores, caractères alphanumériques ; format de la date selon la norme ISO AAAAMMJJ

-Nombres toujours sur au moins 2 unités (en fonction du nombre de fichiers concernés: 01, 002, 0003 etc.). Le même nombre en tout cas

-Élément le plus important en premier pour faciliter la recherche du document

-Terminer par mention de version du type VP/Vdef, V0, 1, 2

-NomProjet_ObjetoDuDocument_Date ou Version

MonProjet_CTreunion_20130211.pdf

Les données collectées sont sauvegardées sur le NAS inserm de notre institut. Le NAS IRMB est en Raid 5 avec Snapshot. Le protocole utilisé est CIFS avec une authentification sur un AD. L'espace de stockage sur le NAS Inserm est nominatif et visible par le groupe "PPC" définis par les administrateurs(Mr Fabrice Galia en local ou la DSI Inserm). Des accès limités et personnalisés peuvent être configurés par l'administrateur. L'ajout ou la suppression d'accès se fait sur demande à l'administrateur.

Les données sont issues des analyseurs avec le format propriétaire et transférer une fois par jour sur le NAS inserm, dans notre espace alloué. Avant archivage, les données pourront être converties en format ouvert.

format ".lcf" généré par le logiciel de pilotage des analyseurs de la société Shimadzu. L'outil "MSconvert" pourra être utilisé. Il supporte la conversion des formats de sortie des analyseurs de masse vers le format standard mzML. mzML est un format ouvert qui pourra être utilisé pour la soumission sur des entrepôts de données.

Origine

- Expérimentation

Différents échantillons biologiques provenant de notre Centre de Ressources Biologiques (plasma, serum, salive, CSF, urine), ou produits par nos collaborateurs (culture de cellules, animaux). Ce sont donc des nouvelles données car ces échantillons n'ont pas encore été analysés.

Les protéines sont extraites de l'échantillon. Les protéines extraites sont ensuite dénaturées, réduites, et alkylées avant une digestion enzymatique (ex: trypsine). Les peptides ainsi générés sont analysés par LC-MRM (TQ 8060NX, Shimadzu). Ces données représentent la "donnée brute". Les métadonnées associées seront documentées.

Les métadonnées seront générées/reportées par l'expérimentateur qui réalise les analyses. C'est cette personne qui sera responsable de ce jeu de métadonnées. Les métadonnées regroupent les informations sur la génération de l'échantillon (informations fournis par le collaborateur), le protocole de préparation de l'échantillon, la partie chromatographie et MS de l'analyse, et les retraitements effectués sur les données brutes.

Les vocabulaires nécessaires à chaque partie seront mis en œuvre. Sur ce site internet ([Ontology Lookup Service < EMBL-EBI](#)) sont répertoriés les vocabulaires contrôlés. Si possible, dans cette liste utilisez les vocabulaires contrôlés National Cancer Institute Thesaurus (NCIt); Experimental Factor Ontology (EFO); PRIDE Proteomics Identifications (PRIDE).

Type de données

- Dataset

Les données brutes sont issues d'expériences et sont générées par les analyseurs sont "la donnée brutes". Elle permet d'identifier/de quantifier les molécules provenant des échantillons. Ce sont des données numériques fixes.

Nature des données

Analyses LC-MS: données chromatographiques et de spectrométrie de masse.
La volumétrie est < 2Mo/ analyse

Format des données

Format ".lcd" généré par le logiciel de pilotage des analyseurs de la société Shimadzu. Avant archivage, les données pourront être converties en format ouvert. L'outil "MSconvert" pourra être utilisé. Il supporte la conversion des formats de sortie des analyseurs de masse vers le format standard mzML. mzML est un format ouvert qui pourra être utilisé pour la soumission sur des entrepôts de données.

Périmètre thématique des données

- Plant Health and Pathology
- Human Nutrition and food security
- Human Health and Pathology
- Animal Health and Pathology

Dosage ELISA (MSD, Simoa, Ella) de molécules

Mode d'obtention des données

- Données générées par la structure

La structure va réaliser des expériences d'identification de molécules sur différentes plateformes technologiques (LC-MS et ELISA). En entrée nous avons des échantillons fournis par un collaborateur interne ou externe à la structure, qui seront préparés selon une procédure, puis analysés. La procédure de préparation de l'échantillon dépend de l'échantillon. Les informations de préparation seront reportées dans les métadatas associées aux datas. La donnée principale, est la donnée de sortie des analyseurs. C'est cette donnée qui doit être conservée, tracée.

Une convention de nommage de l'ensemble des fichiers générés au cours du projet a été adoptée.

Utilisation de règles de nommage standardisées:

-Structuration du nom de fichier : sujet, date, version

-Pas de mots vides, abréviations communément compréhensibles, dans l'idéal, ne pas répéter de termes entre noms du fichier et noms du dossier (même si dans les faits, ce n'est pas si simple...) ; **pas d'espaces** ; **pas de caractères spéciaux** (accentués, symboles... : le nom du fichier doit être lisible par des machines)

-Majuscules, underscores, caractères alphanumériques ; format de la date selon la norme ISO AAAAMMJJ

-Nombres toujours sur au moins 2 unités (en fonction du nombre de fichiers concernés: 01, 002, 0003 etc.). Le même nombre en tout cas

-Élément le plus important en premier pour faciliter la recherche du document

-Terminer par mention de version du type VP/Vdef, V0, 1, 2

-NomProjet_ObjetDuDocument_Date ou Version

MonProjet_CTreunion_20130211.pdf

Les données collectées sont sauvegardées sur le NAS inserm de notre institut. Le NAS IRMB est en Raid 5 avec Snapshot. Le protocole utilisé est CIFS avec une authentification sur un AD. L'espace de stockage sur le NAS Inserm est nominatif et visible par le groupe "PPC" définis par les administrateurs(Mr Fabrice Galia en local ou la DSI Inserm). Des accès limités et personnalisés peuvent être configurés par l'administrateur. L'ajout ou la suppression d'accès se fait sur demande à l'administrateur.

Les données sont issues des analyseurs avec le format propriétaire et transférées une fois par jour sur le NAS inserm, dans notre espace alloué. Avant archivage, les données seront converties en format ouvert.

- ELLA : « Simple Plex Runner » pour lecture plaque et « Simple Plex Explorer » pour lecture des fichiers
- MSD : « Methodical Mind » pour lecture plaque et export. Ouverture des fichiers avec MSD Discovery workbench
- SIMOA : « SIMOA HD-X » pour lecture plaque et résultats bruts avec « Customer Support Tool »

Les documents .txt ou .csv étant des formats ouverts, ils seront éventuellement utilisés pour la soumission sur des entrepôts de données.

Origine

- Expérimentation

Différents échantillons biologiques provenant de notre Centre de Ressources Biologiques (plasma, serum, salive, CSF, urine), ou produits par nos collaborateurs (culture de cellules, animaux). Ce sont donc des nouvelles données car ces échantillons n'ont pas encore été analysés.

L'échantillon est dilué dans le tampon dédié fournis par le kit de dosage. L'échantillon dilué est transféré dans la cuve de lecture pour une mise en contact avec les différents réactifs avant détection de l'absorbance. Les mesures d'absorbances représentent la "donnée brute". Les métadonnées associées seront documentées.

Les métadonnées seront générées/reportées par l'expérimentateur qui réalise les analyses. C'est cette personne qui sera responsable de ce jeu de métadonnées. Les métadonnées regroupent les informations sur la génération de l'échantillon (informations fournis par le collaborateur), le protocole de préparation et d'analyse de l'échantillon, et les retraitements effectués sur les données brutes.

Les vocabulaires nécessaires à chaque partie seront mis en œuvre. Sur ce site internet ([Ontology Lookup Service < EMBL-EBI](#)) sont répertoriés les vocabulaires contrôlés. Si possible, dans cette liste utilisez les vocabulaires contrôlés National Cancer Institute Thesaurus (NCIt); Experimental Factor Ontology (EFO); PRIDE PRoteomics IDentifications (PRIDE).

Type de données

- Dataset

Les données brutes sont issues d'expériences et sont générées par les analyseurs sont "la donnée brutes". Elle permet d'identifier/de quantifier les molécules provenant des échantillons. Ce sont des données numériques fixes.

Nature des données

Analyses Elisa: données d'absorbances

La volumétrie est de:

Simoa : 10 Ko / sample

Le logiciel "Customer Support Tool" est utilisé d'une part en cas de problème technique pour la lecture des données brutes avant export sous format de table, et d'autre part pour la création du Simoa Qualification Test support.

Ainsi, la PPC et la société quanterix peut comprendre et résoudre les problèmes.

Lors de l'envoi du rapport client, tous les problèmes ont été écartés (la série respecte les bonnes performances sur les QCs et la gamme). La donnée brute n'est donc pas sauvegardée.

ELLA, et MSD : 3 Mo / plaque (96 sample)

Format des données

- ELLA : « Simple Plex Runner » pour lecture plaque et « Simple Plex Explorer » pour lecture des fichiers
- MSD : « Methodical Mind » pour lecture plaque et export. Ouverture des fichiers avec MSD Discovery workbench
- SIMOA : « SiMoA HD-X » pour lecture plaque et résultats bruts avec « Customer Support Tool »

Les documents .txt ou .csv étant des formats ouverts, ils seront éventuellement utilisés pour la soumission sur des entrepôts de données.

Périmètre thématique des données

- Human Nutrition and food security
- Human Health and Pathology
- Animal Health and Pathology

Droits de propriété intellectuelle

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

Ouverture des données de recherche | Guide d'analyse du cadre juridique en France. Guide disponible: [Ouvrir la Science - Bibliothèque de la science ouverte](#)

Les décisions d'ouverture des données se prennent au niveau de l'établissement et non pas au niveau de l'agent. La PPC ne prendra donc pas en charge cette responsabilité qui incombera au(x) porteur(s) de projet.

Les droits sur les données et le matériel sont précisés dans l'accord de consortium quand il existe, en relation avec les services juridiques des partenaires du projet. La négociation des termes de l'accord en amont du projet et du plan de gestion est essentielle car elle conditionne en partie les modalités de partage et de diffusion ultérieure des données et du matériel.

S'il n'y a pas d'accord de consortium, préciser les droits dans ce document et prévoir une réunion sur le sujet avec tous les partenaires.

Même dans le cas où les données ne sont pas protégées par un droit (droit d'auteur, droit sui generis du producteur de base de données), il est recommandé de les diffuser accompagnées d'un contrat de licence (CC BY 4.0). Attribuer une **licence** au jeu de données permet de signaler les droits sur le jeu de données et de fixer ses conditions de réutilisation.

Pour de plus ample informations, contactez la "Direction de l'Innovation et des Partenariats" de l'université de Montpellier (dipa@umontpellier.fr)

Confidentialité

Identification des jeux de données contenant des données confidentielles

Lorsque des données à caractère personnel sont produites ou traitées, se mettre en conformité avec le RGPD .

Sur la plateforme, pas de données personnelles ou critiques car les données sont anonymisées par le porteur ou ses collaborateurs avant l'accueil des échantillons sur la plateforme.

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

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 ?

Partage des données

identification de protéines en HRMS

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

La politique de la plateforme est de ne pas interdire ou mettre de restriction au partage des données.

Pour le porteur de projet, se référer à ci-dessous:

Ouverture des données de recherche | Guide d'analyse du cadre juridique en France. Guide disponible: [Ouvrir la Science - Bibliothèque de la science ouverte](#)

Les décisions d'ouverture des données se prennent au niveau de l'établissement et non pas au niveau de l'agent. La PPC ne prendra donc pas en charge cette responsabilité qui incombera au(x) porteur(s) de projet.

Les droits sur les données et le matériel sont précisés dans l'accord de consortium quand il existe, en relation avec les services juridiques des partenaires du projet. La négociation des termes de l'accord en amont du projet et du plan de gestion est essentielle car elle conditionne en partie les modalités de partage et de diffusion ultérieure des données et du matériel.

S'il n'y a pas d'accord de consortium, préciser les droits dans ce document et prévoir une réunion sur le sujet avec tous les partenaires.

Même dans le cas où les données ne sont pas protégées par un droit (droit d'auteur, droit sui generis du producteur de base de données), il est recommandé de les diffuser accompagnées d'un contrat de licence (CC BY 4.0). Attribuer une **licence** au jeu de données permet de signaler les droits sur le jeu de données et de fixer ses conditions de réutilisation.

Pour de plus ample informations, contactez la "Direction de l'Innovation et des Partenariats" de l'université de Montpellier (dipa@umontpellier.fr)

Quelles sont les réutilisations potentielles de ces données ?

En biologie, un maximum de données sont intégrées sur des sites généralistes permettant un accès simple et efficace aux savoirs générés par plusieurs laboratoires de recherche du monde entier (uniprot par exemple).

Ce savoir faire est extrait des publications scientifiques. Grâce aux entrepôts de données et aux métadonnées associées aux données, cette diffusion de savoir va s'en retrouver accélérée.

La réutilisation potentielles de ces données est donc très grande et est une valorisation du travail effectué.

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

format ".baf" généré par le logiciel de pilotage des analyseurs de la société Bruker Daltonics. Avant archivage, les données pourront être convertis en format ouvert. L'outil "MSconvert" pourra être utilisé. Il supporte la conversion des formats de sortie des analyseurs de masse vers le format standard mzML. mzML est un format ouvert qui peut être utiliser pour la soumission sur des entrepôts de données.

Comment les données seront-elles partagées ?

Les zones de stockage utilisées ne permettent pas un accès à des utilisateurs extérieurs.

De plus, les données sont volumineuses et nécessite des logiciels spécifiques (ouverts ou pas) ainsi qu'une expertise pour pouvoir les exploiter.

Pendant le projet, les données brutes ne sont pas partagées et un rapport d'analyse est partagé avec le collaborateur.

En fin de projet, ou pendant le projet, sur demande, les données pourront être partagées sur un dataverse institutionnel du collaborateur (par exemple, Publication sur le portail [Data Inra](#))

Avec qui ?

- Partenaire(s) identifié(s)

Partenaires impliqués dans le projet de recherche

Sous quelle licence ?

- Autre (à préciser dans la zone d'Informations supplémentaires)

CC BY 4.0

dosage de molécules (RNA, métabolites) par des méthodologies basées sur de la LC-MS

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

La politique de la plateforme est de ne pas interdire ou mettre de restriction au partage des données.

Pour le porteur de projet, se référer à ci-dessous:

Ouverture des données de recherche | Guide d'analyse du cadre juridique en France. Guide disponible: [Ouvrir la Science - Bibliothèque de la science ouverte](#)

Les décisions d'ouverture des données se prennent au niveau de l'établissement et non pas au niveau de l'agent. La PPC ne prendra donc pas en charge cette responsabilité qui incombera au(x) porteur(s) de projet.

Les droits sur les données et le matériel sont précisés dans l'accord de consortium quand il existe, en relation avec les services juridiques des partenaires du projet. La négociation des termes de l'accord en amont du projet et du plan de gestion est essentielle car elle conditionne en partie les modalités de partage et de diffusion ultérieure des données et du matériel.

S'il n'y a pas d'accord de consortium, préciser les droits dans ce document et prévoir une réunion sur le sujet avec tous les partenaires.

Même dans le cas où les données ne sont pas protégées par un droit (droit d'auteur, droit sui generis du producteur de base de données), il est recommandé de les diffuser accompagnées d'un contrat de licence (CC BY 4.0). Attribuer une **licence** au jeu de données permet de signaler les droits sur le jeu de données et de fixer ses conditions de réutilisation.

Pour de plus ample informations, contactez la "Direction de l'Innovation et des Partenariats" de l'université de Montpellier (dipa@umontpellier.fr)

Quelles sont les réutilisations potentielles de ces données ?

En biologie, un maximum de données sont intégrées sur des sites généralistes permettant un accès simple et efficace aux savoirs générés par plusieurs laboratoires de recherche du monde entier (uniprot par exemple).

Ce savoir faire est extrait des publications scientifiques. Grâce aux entrepôts de données et aux métadonnées associées aux données, cette diffusion de savoir va s'en retrouver accélérée.

La réutilisation potentielles de ces données est donc très grande et est une valorisation du travail effectué.

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

Format ".lcd" généré par le logiciel de pilotage des analyseurs de la société Shimadzu. Avant archivage, les données pourront être converties en format ouvert. L'outil "MSconvert" pourra être utilisé. Il supporte la conversion des formats de sortie des analyseurs de masse vers le format standard mzML. mzML est un format ouvert qui pourra être utilisé pour la soumission sur des entrepôts de données.

Comment les données seront-elles partagées ?

Les zones de stockage utilisées ne permettent pas un accès à des utilisateurs extérieurs.

De plus, les données sont volumineuses et nécessitent des logiciels spécifiques (ouverts ou pas) ainsi qu'une expertise pour pouvoir les exploiter.

Pendant le projet, les données brutes ne sont pas partagées et un rapport d'analyse est partagé avec le collaborateur.

En fin de projet, ou pendant le projet, sur demande, les données pourront être partagées sur un dataverse institutionnel du collaborateur (par exemple, Publication sur le portail [Data Inra](#))

Avec qui ?

- Partenaire(s) identifié(s)

Partenaires impliqués dans le projet de recherche

Sous quelle licence ?

- Autre (à préciser dans la zone d'Informations supplémentaires)

CC BY 4.0

Dosage de protéines par MRM

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

La politique de la plateforme est de ne pas interdire ou mettre de restriction au partage des données.

Pour le porteur de projet, se référer à ci-dessous:

Ouverture des données de recherche | Guide d'analyse du cadre juridique en France. Guide disponible: [Ouvrir la Science - Bibliothèque de la science ouverte](#)

Les décisions d'ouverture des données se prennent au niveau de l'établissement et non pas au niveau de l'agent. La PPC ne prendra donc pas en charge cette responsabilité qui incombera au(x) porteur(s) de projet.

Les droits sur les données et le matériel sont précisés dans l'accord de consortium quand il existe, en relation avec les services juridiques des partenaires du projet. La négociation des termes de l'accord en amont du projet et du plan de gestion est essentielle car elle conditionne en partie les modalités de partage et de diffusion ultérieure des données et du matériel.

S'il n'y a pas d'accord de consortium, préciser les droits dans ce document et prévoir une réunion sur le sujet avec tous les partenaires.

Même dans le cas où les données ne sont pas protégées par un droit (droit d'auteur, droit sui generis du producteur de base de données), il est recommandé de les diffuser accompagnées d'un contrat de licence (CC BY 4.0). Attribuer une **licence** au jeu de données permet de signaler les droits sur le jeu de données et de fixer ses conditions de réutilisation.

Pour de plus ample informations, contactez la "Direction de l'Innovation et des Partenariats" de l'université de Montpellier (dipa@umontpellier.fr)

Quelles sont les réutilisations potentielles de ces données ?

En biologie, un maximum de données sont intégrées sur des sites généralistes permettant un accès simple et efficace aux savoirs générés par plusieurs laboratoires de recherche du monde entier (uniprot par exemple).

Ce savoir faire est extrait des publications scientifiques. Grâce aux entrepôts de données et aux métadonnées associées aux données, cette diffusion de savoir va s'en retrouver accélérée.

La réutilisation potentielles de ces données est donc très grande et est une valorisation du travail effectué.

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

Format ".lcd" généré par le logiciel de pilotage des analyseurs de la société Shimadzu. Avant archivage, les données pourront être converties en format ouvert. L'outil "MSconvert" pourra être utilisé. Il supporte la conversion des formats de sortie des analyseurs de masse vers le format standard mzML. mzML est un format ouvert qui pourra être utilisé pour la soumission sur des entrepôts de données.

Comment les données seront-elles partagées ?

Les zones de stockage utilisées ne permettent pas un accès à des utilisateurs extérieurs.

De plus, les données sont volumineuses et nécessitent des logiciels spécifiques (ouverts ou pas) ainsi qu'une expertise pour pouvoir les exploiter.

Pendant le projet, les données brutes ne sont pas partagées et un rapport d'analyse est partagé avec le collaborateur.

En fin de projet, ou pendant le projet, sur demande, les données pourront être partagées sur un dataverse institutionnel du collaborateur (par exemple, Publication sur le portail [Data Inra](#))

Avec qui ?

- Partenaire(s) identifié(s)

Partenaires impliqués dans le projet de recherche

Sous quelle licence ?

- Autre (à préciser dans la zone d'Informations supplémentaires)

CC BY 4.0

Dosage ELISA (MSD, Simoa, Ella) de molécules

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

La politique de la plateforme est de ne pas interdire ou mettre de restriction au partage des données.

Pour le porteur de projet, se référer à ci-dessous:

Ouverture des données de recherche | Guide d'analyse du cadre juridique en France. Guide disponible: [Ouvrir la Science - Bibliothèque de la science ouverte](#)

Les décisions d'ouverture des données se prennent au niveau de l'établissement

et non pas au niveau de l'agent. La PPC ne prendra donc pas en charge cette responsabilité qui incombera au(x) porteur(s) de projet.

Les droits sur les données et le matériel sont précisés dans l'accord de consortium quand il existe, en relation avec les services juridiques des partenaires du projet. La négociation des termes de l'accord en amont du projet et du plan de gestion est essentielle car elle conditionne en partie les modalités de partage et de diffusion ultérieure des données et du matériel.

S'il n'y a pas d'accord de consortium, préciser les droits dans ce document et prévoir une réunion sur le sujet avec tous les partenaires.

Même dans le cas où les données ne sont pas protégées par un droit (droit d'auteur, droit sui generis du producteur de base de données), il est recommandé de les diffuser accompagnées d'un contrat de licence (CC BY 4.0). Attribuer une **licence** au jeu de données permet de signaler les droits sur le jeu de données et de fixer ses conditions de réutilisation.

Pour de plus ample informations, contactez la "Direction de l'Innovation et des Partenariats" de l'université de Montpellier (dipa@umontpellier.fr)

Quelles sont les réutilisations potentielles de ces données ?

En biologie, un maximum de données sont intégrées sur des sites généralistes permettant un accès simple et efficace aux savoirs générés par plusieurs laboratoires de recherche du monde entier (uniprot par exemple).

Ce savoir faire est extrait des publications scientifiques. Grâce aux entrepôts de données et aux métadonnées associées aux données, cette diffusion de savoir va s'en retrouver accélérée.

La réutilisation potentielles de ces données est donc très grande et est une valorisation du travail effectué.

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

Les données sont exportées dans des documents .txt ou .csv étant des formats ouverts

Comment les données seront-elles partagées ?

Les zones de stockage utilisées ne permettent pas un accès à des utilisateurs extérieurs.

De plus, les données sont volumineuses et nécessite des logiciels spécifiques (ouverts ou pas) ainsi qu'une expertise pour pouvoir les exploiter.

Pendant le projet, les données brutes ne sont pas partagées et un rapport d'analyse est partagé avec le collaborateur.

En fin de projet, ou pendant le projet, sur demande, les données pourront être partagées sur un dataverse institutionnel du collaborateur (par exemple, Publication sur le portail [Data Inra](#))

Avec qui ?

- Partenaire(s) identifié(s)

Partenaires impliqués dans le projet de recherche

Sous quelle licence ?

- Autre (à préciser dans la zone d'Informations supplémentaires)

CC BY 4.0

Organisation et documentation des données

identification de protéines en HRMS

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

format ".baf" généré par le logiciel de pilotage des analyseurs de la société Bruker Daltonics.

mzML est un format ouvert qui pourra être utilisé pour le dépôt sur des entrepôts de données. L'outil "MSconvert" sera utilisé. Il supporte la conversion des formats de sortie des analyseurs de masse vers le format standard mzML.

Les données générées ont vocation à être diffusées à nos collaborateurs et financeurs. Pendant la durée du projet elles seront sauvegardées.

Le partage des données avec les collaborateurs et financeurs pendant et après la durée du projet sera faite sur demande, sur un dataverse institutionnel du collaborateur (par exemple, Publication sur le portail [Data Inra](#)).

L'archivage est à la discrétion du porteur de projet.

Les seules données qui seront exclues de ce processus sont celles qui n'ont pas satisfait le processus de contrôle qualité. Ce contrôle est effectué par le personnel formés de la plateforme (un tableau de compétences est mis en place dans la norme ISO 9001:2015).

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 ?

Les métadonnées seront générées/reportées par l'expérimentateur qui réalise les analyses. C'est cette personne qui sera responsable de ce jeu de métadonnées. Les métadonnées regroupent les informations sur la génération de l'échantillon (informations fournis par le collaborateur), le protocole de préparation de l'échantillon, la partie chromatographie et MS de l'analyse, et les traitements effectués sur les données brutes.

Les vocabulaires nécessaires à chaque partie seront mis en œuvre.

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)
Life Sciences metadata	saisie via l'interface web	Ontology Lookup Service < EMBL-EBI	Quand nécessaire
Medical metadata	saisie via l'interface web	Ontology Lookup Service < EMBL-EBI	Quand nécessaire
Analytical chemistry metadata	saisie via l'interface web	Ontology Lookup Service < EMBL-EBI	Quand nécessaire

Sur ce site internet sont répertoriés les vocabulaires contrôlés. Si possible, dans cette liste utilisez les vocabulaires contrôlés National Cancer Institute Thesaurus (NCIt); Experimental Factor Ontology (EFO); PRIDE PRoteomics IDentifications (PRIDE).

Le fichier .csv metadata sera stocké avec la data. Il est très complet et bien formaté. A moins d'un changement de règles des entrepôts de dépôt de données spécialisés, des mises à jour ne seront pas nécessaires.

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

non car le standard de métadonnées utilisé est suffisamment complet.

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

Utilisation de règles de nommage standardisées:

- Structuration du nom de fichier : sujet, date, version
- Pas de mots vides, abréviations communément compréhensibles, dans l'idéal, ne pas répéter de termes entre noms du fichier et noms du dossier (même si dans les faits, ce n'est pas si simple...) ; **pas d'espaces** ; **pas de caractères spéciaux** (accentués, symboles... : le nom du fichier doit être lisible par des machines)
- Majuscules, underscores, caractères alphanumériques ; format de la date selon la norme ISO AAAAMMJJ
- Nombres toujours sur au moins 2 unités (en fonction du nombre de fichiers concernés: 01, 002, 0003 etc.). Le même nombre en tout cas
- Élément le plus important en premier pour faciliter la recherche du document
- Terminer par mention de version du type VP/Vdef, V0, 1, 2
- NomProjet_ObjetDuDocument_Date ou Version
MonProjet_CTreunion_20130211.pdf

Organisation des fichiers avec le nom du projet suivi de:

- MiseEnPlace: contenant tous les documents de mise en place du projet entre la PPC et les collaborateurs comme les devis, contrats, annexes scientifiques ...
- SuiviDesEtapas: Etapes/Experiences listées ici par ordre chronologiques contenant les données brutes (ou un chemin d'accès) et les métadatas associées
- SuiviRessources: point budgétaire avec entrées (PP#, UIC, contrat) et dépenses (commandes, salaires), bilan financier
- Valorisation: revue de projet, échange de résultats, publication

Quel est le processus de contrôle qualité des données ?

Les équipements sont utilisés sur une plateforme certifiée ISO 9001:2015 dont les processus de recherche et expertises sont accrédités.

Avant chaque utilisation des instruments de mesure, des contrôles systématiques sont effectués par le personnel formés de la plateforme (un tableau de compétences est mis en place dans la norme ISO 9001:2015): contrôle du niveau des poubelles, contrôle des niveaux de solvants, présence de bulles dans les tubulures, pression de gaz sur les manomètres.

Des contrôles bi-mensuels des instrumentations sont effectuées pour contrôler la calibration et les performances des systèmes analytiques.

dosage de molécules (RNA, métabolites) par des méthodologies basées sur de la LC-MS

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

format ".lcd" généré par le logiciel de pilotage des analyseurs de la société Shimadzu.

mzML est un format ouvert qui pourra être utilisé pour le dépôt sur des entrepôts de données. L'outil "MSconvert" sera utilisé. Il supporte la conversion des formats de sortie des analyseurs de masse vers le format standard mzML.

Les données générées ont vocation à être diffusées à nos collaborateurs et financeurs. Pendant la durée du projet elles seront sauvegardées.

Le partage des données avec les collaborateurs et financeurs pendant et après la durée du projet sera faite sur demande, sur un dataverse institutionnel du collaborateur (par exemple, Publication sur le portail [Data Inra](#)).

L'archivage est à la discrétion du porteur de projet.

Les seules données qui seront exclues de ce processus sont celles qui n'ont pas satisfait le processus de contrôle qualité. Ce contrôle est effectué par le personnel formés de la plateforme (un tableau de compétences est mis en place dans la norme ISO 9001:2015).

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 ?

Les métadonnées seront générées/reportées par l'expérimentateur qui réalise les analyses. C'est cette personne qui sera responsable de ce jeu de métadonnées. Les métadonnées regroupent les informations sur la génération de l'échantillon (informations fournis par le collaborateur), le protocole de préparation de l'échantillon, la partie chromatographie et MS de l'analyse, et les traitements effectués sur les données brutes.

Les vocabulaires nécessaires à chaque partie seront mis en œuvre.

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)
Life Sciences metadata	saisie via l'interface web	Ontology Lookup Service < EMBL-EBI	Quand nécessaire
Medical metadata	saisie via l'interface web	Ontology Lookup Service < EMBL-EBI	Quand nécessaire
Analytical chemistry metadata	saisie via l'interface web	Ontology Lookup Service < EMBL-EBI	Quand nécessaire

Sur ce site internet sont répertoriés les vocabulaires contrôlés. Si possible, dans cette liste utilisez les vocabulaires contrôlés National Cancer Institute Thesaurus (NCIt); Experimental Factor Ontology (EFO); PRIDE PRoteomics IDentifications (PRIDE).

Le fichier .csv metadata sera stocké avec la data. Il est très complet et bien formaté. A moins d'un changement de règles des entrepôts de dépôt de données spécialisés, des mises à jour ne seront pas nécessaires.

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

non car le standard de métadonnées utilisé est suffisamment complet.

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

Utilisation de règles de nommage standardisées:

- Structuration du nom de fichier : sujet, date, version
- Pas de mots vides, abréviations communément compréhensibles, dans l'idéal, ne pas répéter de termes entre noms du fichier et noms du dossier (même si dans les faits, ce n'est pas si simple...) ; **pas d'espaces** ; **pas de caractères spéciaux** (accentués, symboles... : le nom du fichier doit être lisible par des machines)
- Majuscules, underscores, caractères alphanumériques ; format de la date selon la norme ISO AAAAMMJJ
- Nombres toujours sur au moins 2 unités (en fonction du nombre de fichiers concernés: 01, 002, 0003 etc.). Le même nombre en tout cas
- Élément le plus important en premier pour faciliter la recherche du document
- Terminer par mention de version du type VP/Vdef, V0, 1, 2
- NomProjet_ObjetDuDocument_Date ou Version
MonProjet_CTreunion_20130211.pdf

Organisation des fichiers avec le nom du projet suivi de:

- MiseEnPlace: contenant tous les documents de mise en place du projet entre la PPC et les collaborateurs comme les devis, contrats, annexes scientifiques ...
- SuiviDesEtapas: Etapes/Experiences listées ici par ordre chronologiques contenant les données brutes (ou un chemin d'accès) et les métadatas associées
- SuiviRessources: point budgétaire avec entrées (PP#, UIC, contrat) et dépenses (commandes, salaires), bilan financier
- Valorisation: revue de projet, échange de résultats, publication

Quel est le processus de contrôle qualité des données ?

Les équipements sont utilisés sur une plateforme certifiée ISO 9001:2015 dont les processus de recherche et expertises sont accrédités.

Avant chaque utilisation des instruments de mesure, des contrôles systématiques sont effectués par le personnel formés de la plateforme (un tableau de compétences est mis en place dans la norme ISO 9001:2015): contrôle du niveau des poubelles, contrôle des niveaux de solvants, présence de bulles dans les tubulures, pression de gaz sur les manomètres.

Des contrôles mensuels des instrumentations sont effectués pour contrôler la calibration et les performances des systèmes analytiques.

Dosage de protéines par MRM

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

format ".lcd" généré par le logiciel de pilotage des analyseurs de la société Shimadzu.

mzML est un format ouvert qui pourra être utilisé pour le dépôt sur des entrepôts de données. L'outil "MSconvert" sera utilisé. Il supporte la conversion des formats de sortie des analyseurs de masse vers le format standard mzML.

Les données générées ont vocation à être diffusées à nos collaborateurs et financeurs. Pendant la durée du projet elles seront sauvegardées.

Le partage des données avec les collaborateurs et financeurs pendant et après la durée du projet sera faite sur demande, sur un dataverse institutionnel du collaborateur (par exemple, Publication sur le portail [Data Inra](#)).

L'archivage est à la discrétion du porteur de projet.

Les seules données qui seront exclus de ce processus sont celles qui n'ont pas satisfait le processus de contrôle qualité. Ce contrôle est effectué par le personnel formés de la plateforme (un tableau de compétences est mis en place dans la norme ISO 9001:2015).

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 ?

Les métadonnées seront générées/reportées par l'expérimentateur qui réalise les analyses. C'est cette personne qui sera responsable de ce jeu de métadonnées. Les métadonnées regroupent les informations sur la génération de l'échantillon (informations fournis par le collaborateur), le protocole de préparation de l'échantillon, la partie chromatographie et MS de l'analyse, et les retraitements effectués sur les données brutes.

Les vocabulaires nécessaires à chaque partie seront mis en oeuvre.

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)
Life Sciences metadata	saisie via l'interface web	Ontology Lookup Service < EMBL-EBI	Quand nécessaire
Medical metadata	saisie via l'interface web	Ontology Lookup Service < EMBL-EBI	Quand nécessaire
Analytical chemistry metadata	saisie via l'interface web	Ontology Lookup Service < EMBL-EBI	Quand nécessaire

Sur ce site internet sont répertoriés les vocabulaires contrôlés. Si possible, dans cette liste utilisez les vocabulaires contrôlés National Cancer Institute Thesaurus (NCIt); Experimental Factor Ontology (EFO); PRIDE PRoteomics IDentifications (PRIDE).

Le fichier .csv metadata sera stocké avec la data. Il est très complet et bien formaté. A moins d'un changement de règles des entrepôts de dépôt de données spécialisés, des mises à jour ne seront pas nécessaires.

Une documentation complémentaire aux métadonnées est-elle nécessaire pour décrire les données et assurer leur

réutilisabilité sur le long terme ?

non car le standard de métadonnées utilisé est suffisamment complet.

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

Utilisation de règles de nommage standardisées:

- Structuration du nom de fichier : sujet, date, version
- Pas de mots vides, abréviations communément compréhensibles, dans l'idéal, ne pas répéter de termes entre noms du fichier et noms du dossier (même si dans les faits, ce n'est pas si simple...) ; **pas d'espaces** ; **pas de caractères spéciaux** (accentués, symboles... : le nom du fichier doit être lisible par des machines)
- Majuscules, underscores, caractères alphanumériques ; format de la date selon la norme ISO AAAAMMJJ
- Nombres toujours sur au moins 2 unités (en fonction du nombre de fichiers concernés: 01, 002, 0003 etc.). Le même nombre en tout cas
- Élément le plus important en premier pour faciliter la recherche du document
- Terminer par mention de version du type VP/Vdef, V0, 1, 2
- NomProjet_ObjetDuDocument_Date ou Version
MonProjet_CTreunion_20130211.pdf

Organisation des fichiers avec le nom du projet suivi de:

- MiseEnPlace: contenant tous les documents de mise en place du projet entre la PPC et les collaborateurs comme les devis, contrats, annexes scientifiques ...
- SuiviDesEtapes: Etapes/Experiences listées ici par ordre chronologiques contenant les données brutes (ou un chemin d'accès) et les métadatas associées
- SuiviRessources: point budgétaire avec entrées (PP#, UIC, contrat) et dépenses (commandes, salaires), bilan financier
- Valorisation: revue de projet, échange de résultats, publication

Quel est le processus de contrôle qualité des données ?

Les équipements sont utilisés sur une plateforme certifiée ISO 9001:2015 dont les processus de recherche et expertises sont accrédités.

Avant chaque utilisation des instruments de mesure, des contrôles systématiques sont effectués par le personnel formés de la plateforme (un tableau de compétences est mis en place dans la norme ISO 9001:2015): contrôle du niveau des poubelles, contrôle des niveaux de solvants, présence de bulles dans les tubulures, pression de gaz sur les manomètres.

Des contrôles mensuels des instrumentations sont effectués pour contrôler la calibration et les performances des systèmes analytiques.

Dosage ELISA (MSD, Simoa, Ella) de molécules

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

- ELLA : « Simple Plex Runner » pour lecture plaque et « Simple Plex Explorer » pour lecture des fichiers
- MSD : « Methodical Mind » pour lecture plaque et export. Ouverture des fichiers avec MSD Discovery workbench
- SIMOA : « SiMoA HD-X » pour lecture plaque et résultats bruts avec « Customer Support Tool »

Les documents .txt ou .csv étant des formats ouverts, ils pourront être utilisés pour le dépôt sur des entrepôts de données.

Les données générées ont vocation à être diffusées à nos collaborateurs et financeurs. Pendant la durée du projet elles seront sauvegardées.

Le partage des données avec les collaborateurs et financeurs pendant et après la durée du projet sera faite sur demande, sur un dataverse institutionnel du collaborateur (par exemple, Publication sur le portail [Data Inra](#)).

L'archivage est à la discrétion du porteur de projet.

Les seules données qui seront exclues de ce processus sont celles qui n'ont pas satisfait le processus de contrôle qualité. Ce contrôle est effectué par le personnel formés de la plateforme (un tableau de compétences est mis en place dans la norme ISO 9001:2015).

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 ?

Les métadonnées seront générées/reportées par l'expérimentateur qui réalise les analyses. C'est cette personne qui sera responsable de ce jeu de métadonnées. Les métadonnées regroupent les informations sur la génération de l'échantillon (informations fournis par le collaborateur), le protocole de préparation et d'analyse de l'échantillon, et les retraitements effectués sur les données brutes.

Les vocabulaires nécessaires à chaque partie seront mis en oeuvre.

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)
Life Sciences metadata	saisie via l'interface web	Ontology Lookup Service < EMBL-EBI	Quand nécessaire
Medical metadata	saisie via l'interface web	Ontology Lookup Service < EMBL-EBI	Quand nécessaire
Analytical chemistry metadata	saisie via l'interface web	Ontology Lookup Service < EMBL-EBI	Quand nécessaire

Sur ce site internet sont répertoriés les vocabulaires contrôlés. Si possible, dans cette liste utilisez les vocabulaires contrôlés National Cancer Institute Thesaurus (NCIt); Experimental Factor Ontology (EFO); PRIDE PRoteomics IDentifications (PRIDE).

Le fichier .csv metadata sera stocké avec la data. Il est très complet et bien formaté. A moins d'un changement de règles des entrepôts de dépôt de données spécialisés, des mises à jour ne seront pas nécessaires.

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

non car le standard de métadonnées utilisé est suffisamment complet.

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

Utilisation de règles de nommage standardisées:

- Structuration du nom de fichier : sujet, date, version
 - Pas de mots vides, abréviations communément compréhensibles, dans l'idéal, ne pas répéter de termes entre noms du fichier et noms du dossier (même si dans les faits, ce n'est pas si simple...) ; **pas d'espaces** ; **pas de caractères spéciaux** (accentués, symboles... : le nom du fichier doit être lisible par des machines)
 - Majuscules, underscores, caractères alphanumériques ; format de la date selon la norme ISO AAAAMMJJ
 - Nombres toujours sur au moins 2 unités (en fonction du nombre de fichiers concernés: 01, 002, 0003 etc.). Le même nombre en tout cas
 - Élément le plus important en premier pour faciliter la recherche du document
 - Terminer par mention de version du type VP/Vdef, V0, 1, 2
 - NomProjet_ObjetDuDocument_Date ou Version
- MonProjet_CTreunion_20130211.pdf

Organisation des fichiers avec le nom du projet suivi de:

- MiseEnPlace: contenant tous les documents de mise en place du projet entre la PPC et les collaborateurs comme les devis, contrats, annexes scientifiques ...
- SuiviDesEtapes: Etapes/Expériences listées ici par ordre chronologiques contenant les données brutes (ou un chemin d'accès) et

les métadatas associées

- SuiviRessources: point budgétaire avec entrées (PP#, UIC, contrat) et dépenses (commandes, salaires), bilan financier
- Valorisation: revue de projet, échange de résultats, publication

Quel est le processus de contrôle qualité des données ?

Les équipements sont utilisés sur une plateforme certifiée ISO 9001:2015 dont les processus de recherche et expertises sont accrédités.

Avant chaque utilisation des instruments de mesure, des contrôles systématiques sont effectués par le personnel formés de la plateforme (un tableau de compétences est mis en place dans la norme ISO 9001:2015): contrôle du niveau des poubelles, contrôle des niveaux des buffers, présence de bulles dans les tubulures.

Des échantillons de contrôles qualité et des gammes sont inclus dans chaque série de mesure (chaque kit).

Stockage et sécurité des données

identification de protéines en HRMS

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

Les données générées sont automatiquement transférées sur le NAS IRMB (Stockage interne à l'institut pour éviter des ouvertures vers l'extérieur - Gestionnaire Inserm) toute les nuits avec un logiciel SynBackFree.

Le NAS IRMB est en Raid 5 avec Snapshot. Le protocole utilisé est CIFS avec une authentification sur un AD

Quelle est la volumétrie actuelle et prévisionnelle ?

La volumétrie est < 3Go/ analyse

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...*

Les données seront stockées sur le data center de mesoLR au CINES de Montpellier ([Plateformes HPC et Données – MESO@LR \(umontpellier.fr\)](#)).

L'accès se fait par un mode de partage selon le protocole CIFS.

A cela s'ajoutent des services de réplication de données et de snapshots à la fois complémentaires et offrant une excellente protection contre les ransomwares.

La réplication est une copie des données dans une salle physiquement distante du stockage initial au CINES. La copie étant complète, le volume des données est donc doublé. Dans ce cas, la réplication se fait aussi au cines mais dans une salle distante (deux salles reliées par un long couloir avec porte coupe-feu).

Les snapshots sont des copies (sortes de photographies) des données régulières et prédéfinies (fréquence réglée à 5 fois par 24h), effectuées sur le même espace disque qui permettent de revenir en arrière sur la ou les versions précédentes. Les snapshots ne stockent que les différences sur les fichiers modifiés depuis le dernier snapshot, l'impact en termes de volume dépend des modifications effectuées entre les snapshots.

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 ?

L'espace de stockage sur le NAS Inserm est nominatif et visible par le groupe définis par les administrateurs (Mr Fabrice Galia en local ou la DSI Inserm). Des accès limités et personnalisés peuvent être configurés par l'administrateur. L'ajout ou la suppression d'accès se fait sur demande à l'administrateur.

Les zones de stockage utilisées ne permettent pas un accès à des utilisateurs extérieurs.

De plus, les données sont volumineuses et nécessite des logiciels spécifiques (ouverts ou pas) ainsi qu'une expertise pour pouvoir les exploiter.

Pendant le projet, les données brutes ne sont pas partagées et un rapport d'analyse est partagé avec le collaborateur.

En fin de projet, ou pendant le projet, sur demande, les données pourront être partagées sur un dataverse institutionnel du collaborateur (par exemple, Publication sur le portail [Data Inra](#))

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

L'environnement de travail se situe dans l'espace inserm qui gère les accès (pare-feu), la sécurité et la protection de nos ordinateurs et donc de nos données brutes et tous les fichiers qui en découlent. Un antivirus géré par la tutelle est installé sur chaque ordinateur.

dosage de molécules (RNA, métabolites) par des méthodologies basées sur de la LC-MS

Quels sont les types de flux empruntés par les données et les supports utilisés pour les stocker ?

(Faire éventuellement un lien vers un schéma)

Les données générées sont automatiquement transférées sur le NAS IRMB (Stockage interne à l'institut pour éviter des ouvertures vers l'extérieur - Gestionnaire Inserm) toute les nuits avec un logiciel SynBackFree.

Le NAS IRMB est en Raid 5 avec Snapshot. Le protocole utilisé est CIFS avec une authentification sur un AD

Quelle est la volumétrie actuelle et prévisionnelle ?

La volumétrie est < 2Mo/ analyse

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...*

Les données seront stockées sur le data center de mesoLR au CINES de Montpellier ([Plateformes HPC et Données – MESO@LR \(umontpellier.fr\)](#)).

L'accès se fait par un mode de partage selon le protocole CIFS.

A cela s'ajoutent des services de réplication de données et de snapshots à la fois complémentaires et offrant une excellente protection contre les ransomwares.

La réplication est une copie des données dans une salle physiquement distante du stockage initial au CINES. La copie étant complète, le volume des données est donc doublé. Dans ce cas, la réplication se fait aussi au cines mais dans une salle distante (deux salles reliées par un long couloir avec porte coupe-feu).

Les snapshots sont des copies (sortes de photographies) des données régulières et prédéfinies (fréquence réglée à 5 fois par 24h), effectuées sur le même espace disque qui permettent de revenir en arrière sur la ou les versions précédentes. Les snapshots ne stockent que les différences sur les fichiers modifiés depuis le dernier snapshot, l'impact en termes de volume dépend des modifications effectuées entre les snapshots.

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 ?

L'espace de stockage sur le NAS Inserm est nominatif et visible par le groupe définis par les administrateurs (Mr Fabrice Galia en local ou la DSI Inserm). Des accès limités et personnalisés peuvent être configurés par l'administrateur. L'ajout ou la suppression d'accès se fait sur demande à l'administrateur.

Les zones de stockage utilisées ne permettent pas un accès à des utilisateurs extérieurs.

De plus, les données sont volumineuses et nécessitent des logiciels spécifiques (ouverts ou pas) ainsi qu'une expertise pour pouvoir les exploiter.

Pendant le projet, les données brutes ne sont pas partagées et un rapport d'analyse est partagé avec le collaborateur.

En fin de projet, ou pendant le projet, sur demande, les données pourront être partagées sur un dataverse institutionnel du collaborateur (par exemple, Publication sur le portail [Data Inra](#))

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

L'environnement de travail se situe dans l'espace inserm qui gère les accès (pare-feu), la sécurité et la protection de nos ordinateurs et donc de nos données brutes et tous les fichiers qui en découlent. Un antivirus géré par la tutelle est installé sur chaque ordinateur.

Dosage de protéines par MRM

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

Les données générées sont automatiquement transférées sur le NAS IRMB (Stockage interne à l'institut pour éviter des ouvertures vers l'extérieur - Gestionnaire Inserm) toute les nuits avec un logiciel SynBackFree.

Le NAS IRMB est en Raid 5 avec Snapshot. Le protocole utilisé est CIFS avec une authentification sur un AD

Quelle est la volumétrie actuelle et prévisionnelle ?

La volumétrie est < 2Mo/ analyse

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...

Les données seront stockées sur le data center de mesoLR au CINES de Montpellier ([Plateformes HPC et Données – MESO@LR \(umontpellier.fr\)](#)).

L'accès se fait par un mode de partage selon le protocole CIFS.

A cela s'ajoutent des services de réplication de données et de snapshots à la fois complémentaires et offrant une excellente protection contre les ransomwares.

La réplication est une copie des données dans une salle physiquement distante du stockage initial au CINES. La copie étant complète, le volume des données est donc doublé. Dans ce cas, la réplication se fait aussi au cines mais dans une salle distante (deux salles reliées par un long couloir avec porte coupe-feu).

Les snapshots sont des copies (sortes de photographies) des données régulières et prédéfinies (fréquence réglée à 5 fois par 24h), effectuées sur le même espace disque qui permettent de revenir en arrière sur la ou les versions précédentes. Les snapshots ne stockent que les différences sur les fichiers modifiés depuis le dernier snapshot, l'impact en termes de volume dépend des modifications effectuées entre les snapshots.

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 ?

L'espace de stockage sur le NAS Inserm est nominatif et visible par le groupe définis par les administrateurs (Mr Fabrice Galia en local ou la DSI Inserm). Des accès limités et personnalisés peuvent être configurés par l'administrateur. L'ajout ou la suppression d'accès se fait sur demande à l'administrateur.

Les zones de stockage utilisées ne permettent pas un accès à des utilisateurs extérieurs.

De plus, les données sont volumineuses et nécessitent des logiciels spécifiques (ouverts ou pas) ainsi qu'une expertise pour pouvoir les exploiter.

Pendant le projet, les données brutes ne sont pas partagées et un rapport d'analyse est partagé avec le collaborateur.

En fin de projet, ou pendant le projet, sur demande, les données pourront être partagées sur un dataverse institutionnel du collaborateur (par exemple, Publication sur le portail [Data Inra](#))

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

L'environnement de travail se situe dans l'espace inserm qui gère les accès (pare-feu), la sécurité et la protection de nos ordinateurs et donc de nos données brutes et tous les fichiers qui en découlent. Un antivirus géré par la tutelle est installé sur chaque ordinateur.

Dosage ELISA (MSD, Simoa, Ella) de molécules

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

Les données générées sont automatiquement transférées sur le NAS IRMB (Stockage interne à l'institut pour éviter des ouvertures vers l'extérieur - Gestionnaire Inserm) toute les nuits avec un logiciel SynBackFree.

Le NAS IRMB est en Raid 5 avec Snapshot. Le protocole utilisé est CIFS avec une authentification sur un AD

Quelle est la volumétrie actuelle et prévisionnelle ?

La volumétrie est de:

Simoa : 10 Ko / sample

Ella, et MSD : 3 Mo / plaque (96 sample)

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...

Les données seront stockées sur le data center de mesoLR au CINES de Montpellier ([Plateformes HPC et Données – MESO@LR \(umontpellier.fr\)](#)).

L'accès se fait par un mode de partage selon le protocole CIFS.

A cela s'ajoutent des services de réplication de données et de snapshots à la fois complémentaires et offrant une excellente protection contre les ransomwares.

La réplication est une copie des données dans une salle physiquement distante du stockage initial au CINES. La copie étant complète, le volume des données est donc doublé. Dans ce cas, la réplication se fait aussi au cines mais dans une salle distante (deux salles reliées par un long couloir avec porte coupe-feu).

Les snapshots sont des copies (sortes de photographies) des données régulières et prédéfinies (fréquence réglée à 5 fois par 24h), effectuées sur le même espace disque qui permettent de revenir en arrière sur la ou les versions précédentes. Les snapshots

ne stockant que les différences sur les fichiers modifiés depuis le dernier snapshot, l'impact en termes de volume dépend des modifications effectuées entre les snapshots.

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 ?

L'espace de stockage sur le NAS Inserm est nominatif et visible par le groupe définis par les administrateurs (Mr Fabrice Galia en local ou la DSI Inserm). Des accès limités et personnalisés peuvent être configurés par l'administrateur. L'ajout ou la suppression d'accès se fait sur demande à l'administrateur.

Les zones de stockage utilisées ne permettent pas un accès à des utilisateurs extérieurs.

De plus, les données sont volumineuses et nécessite des logiciels spécifiques (ouverts ou pas) ainsi qu'une expertise pour pouvoir les exploiter.

Pendant le projet, les données brutes ne sont pas partagées et un rapport d'analyse est partagé avec le collaborateur.

En fin de projet, ou pendant le projet, sur demande, les données pourront être partagées sur un dataverse institutionnel du collaborateur (par exemple, Publication sur le portail [Data Inra](#))

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

L'environnement de travail se situe dans l'espace inserm qui gère les accès (pare-feu), la sécurité et la protection de nos ordinateurs et donc de nos données brutes et tous les fichiers qui en découlent. Un antivirus géré par la tutelle est installé sur chaque ordinateur.

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 ?

Les données brutes en sortie d'analyseur, associées à leurs métadonnées, sont indispensables à la ré-utilisation des datas. Ce sont ces données qui seront conservées à moyen, long terme.

Les données n'ayant pas satisfait les critères de qualité (problèmes sur la préparation de l'échantillon biologique ou des répliquats biologiques, problèmes lors de la préparation des échantillons, ou problèmes sur les performances des QCs machines) ne seront pas conservées.

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 ?

La PPC ne propose pas de solution d'archivage. Cela reste à la discrétion du porteur de projet.

Quelle est la durée de conservation des données ?

Les données seront sauvegardées à moyen terme sur MesoLR et seront conservées deux ans après la clôture du projet.

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

La PPC ne propose pas de solution d'archivage. Cela reste à la discrétion du porteur de projet.