
DMP de structure, plateforme "Bordeaux Metabolome"

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 de structure, plateforme "Bordeaux Metabolome"
Domaines de recherche (selon classification de l'OCDE)	
Langue	fra
Date de création	2021-12-20
Date de dernière modification	2022-06-29
Licence	Creative Commons Attribution Share Alike 4.0 International
Documents (publications, rapports, brevets, plan expérimental....), sites web associés	<ul style="list-style-type: none">• Site web plateforme Bordeaux Metabolome : https://metabolome.cgfb.u-bordeaux.fr• Page web INRAE pour la plateforme Bordeaux Metabolome : https://www6.bordeaux-aquitaine.inrae.fr/bfp/Ressources/Plateforme-Metabolome

Renseignements sur le projet

Titre du projet	Bordeaux Metabolome
Résumé	PGD de structure, plateforme Bordeaux Metabolome
Sources de financement	<ul style="list-style-type: none">• Agence Nationale de la Recherche : ANR-11-INBS-0010• Agence Nationale de la Recherche : ANR-11-INBS-0012
Partenaires	<ul style="list-style-type: none">• University of Bordeaux ()• National Center for Scientific Research ()• National Research Institute for Agriculture, Food and the Environment ()

Produits de recherche :

1. Pôle Métabolome (Jeu de données)
2. Pôle Lipidome (Jeu de données)
3. Pôle Polyphénols (Jeu de données)
4. Pôle Phénotypage Métabolique à haut débit (plateau HitMe) et Microfluidique (Jeu de données)

Contributeurs

Nom	Affiliation	Rôles
Benaben David - https://orcid.org/0000-0001-9930-9363	UMR 1332 Biologie du Fruit et Pathologie, INRAE, Université de Bordeaux, 33882 Villenave d'Ornon, France	<ul style="list-style-type: none"> • Responsable du plan de gestion de données
Fouillen Laetitia - https://orcid.org/0000-0002-1204-9296	Laboratoire de Biogénèse Membranaire, CNRS, UMR 5200, Université de Bordeaux, 33883 Villenave D'Ornon, France	<ul style="list-style-type: none"> • Personne contact pour les données (Lipidome)
Gibon Yves - https://orcid.org/0000-0001-8161-1089	UMR 1332 Biologie du Fruit et Pathologie, INRAE, Université de Bordeaux, 33882 Villenave d'Ornon, France	<ul style="list-style-type: none"> • Personne contact pour les données (HitMe)
Jacob Daniel - https://orcid.org/0000-0002-6687-7169	UMR 1332 Biologie du Fruit et Pathologie, INRAE, Université de Bordeaux, 33882 Villenave d'Ornon, France	
Petriacq Pierre - https://orcid.org/0000-0001-8151-7420	UMR 1332 Biologie du Fruit et Pathologie, INRAE, Université de Bordeaux, 33882 Villenave d'Ornon, France	<ul style="list-style-type: none"> • Coordinateur du projet • Personne contact pour les données (Métabolome)
Valls-fonayet Josep - https://orcid.org/0000-0002-1359-4114	Université de Bordeaux, UR Œnologie, MIB, EA 4577, USC 1366 INRAE, Bordeaux INP, ISVV, F-33140, Villenave d'Ornon, France	<ul style="list-style-type: none"> • Personne contact pour les données (Polyphénols)
Valls-Fonayet Josep - https://orcid.org/0000-0002-1359-4114	Unité de Recherche Œnologie, Molécules d'Intérêt Biologique, Université de Bordeaux, ISVV, EA 4577, USC 1366 INRA, 210 Chemin de Leysotte, F-33882 Villenave d'Ornon, France	

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 de structure, plateforme "Bordeaux Metabolome"

Informations sur la structure

Nom de la structure

Bordeaux Metabolome

Type de structure

- Plateforme, plateau technique
- ISC (Infrastructure Scientifique Collective)

Label:

- IBISA, Plateforme de Recherche en Sciences du Vivant en 2008
 - INRA, Plateforme Stratégique en 2008
 - INRA, Plateforme Stratégique Nationale en 2013.
 - INRA, Infrastructure Scientifique Commune (ISC) en 2018.
 - Université de Bordeaux, Plateforme de Recherche par l'Université de Bordeaux en 2019
-

Identifiant de la structure

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

DataCite: <https://doi.org/10.15454/1.5572412770331912E12>

Responsabilités dans la structure

Prénom Nom	Courriel	Rôle
Pierre Petriacq	pierre.petriacq@inrae.fr	Responsable de la plateforme
Josep Valls-Fonayet	josep.valls-fonayet@u-bordeaux.fr	Responsable adjoint de la plateforme
Daniel Jacob	jacob.daniel@inrae.fr	Référent gestion des données
David Benaben	david.benaben@inrae.fr	Référent gestion des données

Etablissement(s) tutelle(s)

- INRAE
- Université de Bordeaux
- CNRS

Plateforme portée par trois tutelles (Université de Bordeaux, INRAE et CNRS) via ses trois UMR d'adossment : l'UMR1332 Biologie du Fruit et Pathologie, l'Unité d'œnologie (EA 4577 USC INRAE 1366) et l'UMR5200 Laboratoire de Biogenèse Membranaire.

Département de rattachement INRAE (ou anciens départements Inra)

- BAP : Biologie et amélioration des plantes

Bordeaux Metabolome est l'une des plateformes technologiques du département INRAE BAP et contribue aussi à des projets portés par des acteurs des départements SPE, MICA, PHASE et EA

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

- Instituts (INRAE, Université de Bordeaux, CNRS)
- Contrats de recherche/Privés/Prestations
- Projets (Investissement d'Avenir, IBISA, EPPN 2020, Région Nouvelle-Aquitaine, MetaboHUB ANR-11-INBS-0010, Phenome Emphasis ANR-11-INBS-0012)

Informations sur le plan de gestion

DOI (*version publiée du plan de gestion*)

Question sans réponse.

Historique des versions

Date	n° de version	Status	Auteur	Affiliation de l'auteur (se reporter à l' annuaire INRAE)	Validé par	Validé le
2022-01-01	v1	en cours	David Benaben	INRAE, Biologie du Fruit et Pathologie		

Présentation générale des données

Pôle Métabolome

Mode d'obtention des données

- Données générées par la structure

Phénotypage biochimique d'échantillons végétaux

- Dosages robotisés en microplaques: protéines totales, amidon, saccharose, glucose, fructose, malate, fumarate (sous réserve), citrate, acides aminés libres totaux, glutamate, proline, chlorophylles a et b (à partir d'une extraction)
- Quantification d'acides aminés par UHPLC-fluorimétrie
- Analyse de composés phénoliques (HPLC-UV/Visible, LC-MS ion trap, LC-NMR)

Profils métabolomiques-Métabolome d'échantillons végétaux

- Quantification des métabolites polaires majeurs par RMN-1H d'extraits polaires
- Réalisation et analyse d'empreintes métabolomiques par RMN-1H d'extraits polaires
- Réalisation et analyse d'empreintes métabolomiques par LC-QqTOF-MS d'extraits semi-polaires
- Profils lipidiques-Lipidome d'échantillons végétaux

Analyse de lipides (phospholipides, lipides neutres, sphingolipides...)

- Analyse et quantification des acides gras par GC-FID ou GC-MS
- Analyse et quantification des lipides par GC-FID, GC-MS ou TLC
- Analyse des cires et polymères lipidiques par GC-MS et GC-FID
- Quantification relative de sphingolipides de plantes par LC-MS
- Réalisation et analyse de profils lipidomiques par MS/MS et LC-Qtrap-MS

Identification de composés

- Analyses structurales de composés phénoliques (MS et RMN)

Étude du métabolisme

- Mesures robotisées d'enzymes du métabolisme central (glycolyse, TCA, Calvin-pentoses phosphates, saccharose, assimilation de l'azote)
- Réalisation de spectres RMN pour la mesure d'enrichissements isotopiques C13 sur des fractions, pour la mesure de flux métaboliques

Origine

- Analyse
- Expérimentation

Équipements générant les données:

- LCMS MicrOTOf-Q (Bruker)
- UHPLC-LTQ/Orbitrap (Thermo)
- UHPLC-Fluo-DAD (Dionex)
- Spectromètre RMN 500 MHz AVANCE III (Bruker)
- pH mètre pour extraits pour profils métaboliques ou métabolomiques par RMN (Mettler Toledo)
- Robot Titration BTpH
- Logiciels de traitements de données :
 - NMRProcFlow : <https://nmrprocflow.org>
 - BioStatFlow : <http://biostatflow.org>

Type de données

- Dataset
- Workflow

Dataset : Tableaux de données expérimentales le plus souvent associés à un plan expérimental (DoE)

Workflow : Séquence de traitement de données : ex fichiers de macrocommandes issues de NMRProcFlow, Jupyter notebooks.

Nature des données

Format des données

Format Spectrométrie de Masse (MS): https://en.wikipedia.org/wiki/Mass_spectrometry_data_format

Données brutes, format propriétaires (.raw, .dcl, .pai2, etc.)

Données traitées, format ouverts (mzML, mzXML)

Spectre RMN - Format Bruker (répertoire structuré), Format ouvert (nmrML)

Jupyter notebooks - Format JSON (.ipynb)

Comma Separated Values (.csv, .txt)

Office Open XML (.xlsx, .docx, .pptx)

Rapports (.png, .pdf, .docx, .pptx)

Périmètre thématique des données

- Omics
-

Pôle Lipidome

Mode d'obtention des données

- Données générées par la structure

La plateforme permet d'obtenir l'identification, la caractérisation ou la quantification (absolue ou relative) de divers lipides :

- Quantification par densitométrie des classes de lipides
 - Identification et quantification d'acides gras (de longue chaînes, hydroxylé...), de waxes, de composés de la cutine et de la subérine, des « long chain based » (LCB)
 - Quantification relative des espèces moléculaires de phospholipides (PC, PE, PG, PS, PI and PIPx), lipides neutres (TAG, DAG, steryl ester...), des sphingolipides de plantes (Cer, GluCer et GIPC A&B)
 - Régionalisation pour les espèces moléculaires de TAG
 - Quantification des Alkyl HydroxyCinnamates (alkyl ferulate, alkyl coumarate et alkyl caffeate)
-

Origine

- Analyse
- Expérimentation

Équipements générant les données:

- LC-MS (Sciex)
 - GC-MS (Agilent)
 - GC-FID (Agilent)
 - TLC
 - Logiciels de traitements de données associés aux équipements
-

Type de données

Question sans réponse.

Nature des données

- profils lipidomiques
- identification et analyse structurale de lipides
- analyse fonctionnelle du métabolisme des lipides

Format des données

- LC-MS Sciex : format propriétaire*.wiff ; pouvant être transformé en format ouvert *.mzml
- GC-MS Agilent : Format propriétaire ; pouvant être transformé en format ouvert *.mzml
- GC-FID Agilent : Format propriétaire
- TLC : Image sous format *.TIF
- Logiciels de traitements de données associés aux équipements: fichier excel *.xlsx

Les expériences peuvent être consignées dans une fiche (sous format .xls) résumant l'objectif, le matériel utilisé, les conditions expérimentales, l'emplacement des données brutes et l'analyse des résultats obtenus.

Périmètre thématique des données

- Omics

Nnombreux organismes, allant des plantes, aux animaux en passant par les levures.

Pôle Polyphénols

Mode d'obtention des données

- Données générées par la structure
- Analysis of phenolics (HPLC-UV/Visible, LC-MS ion trap, LC-NMR)
- Identification of compounds using LC-SPE-NMR and multidimensional NMR,

Origine

- Analyse
- Expérimentation

Équipements générant les données:

- HPLC-UV/Visible
- LC-MS ion trap
- LC-NMR
- Logiciels de traitements de données associés aux équipements

Type de données

- Dataset

* Tableaux de données de quantification (métaboliques)

Nature des données

- profils métabolomiques
- identification et analyse phénolique

Format des données

Format Spectrométrie de Masse (MS): https://en.wikipedia.org/wiki/Mass_spectrometry_data_format

Données brutes, format propriétaires (.raw, .dcl, .pai2, etc.)

Données traitées, format ouverts (mzML, mzXML)

Spectre RMN - Format Bruker (répertoire structuré), Format ouvert (nmrML)

Comma Separated Values (.csv, .txt)

Office Open XML (.xlsx, .docx, .pptx)

Rapports (.png, .pdf, .docx, .pptx)

Périmètre thématique des données

- Omics

Essentiellement analyses phénotypiques, phénoliques et métaboliques de vins de différentes régions (Bordeaux, Bourgogne, ...)

Pôle Phénotypage Métabolique à haut débit (plateau HitMe) et Microfluidique

Mode d'obtention des données

- Données générées par la structure

Dosages robotisés en microplaques: protéines totales, amidon, saccharose, glucose, fructose, malate, fumarate (sous réserve), citrate, acides aminés libres totaux, glutamate, proline, chlorophylles a et b (à partir d'une extraction)

Mesures robotisées d'enzymes du métabolisme central (glycolyse, TCA, Calvin-pentoses phosphates, saccharose, assimilation de l'azote)

Origine

- Expérimentation

Équipements générant les données:

- Robots de pipetage Starlet (Hamilton)
- Lecteurs de microplaques UV/visible (Safas)
- Lecteur de microplaques multifonction Fluo Xenius (Safas)

Type de données

- Dataset

Dataset/Tableau de données

Nature des données

Analyses métabolomiques

Format des données

Comma Separated Values (.csv)

Office Open XML (.xlsx)

Données brutes, fichiers issus du spectrophotomètre, format propriétaire (.amp96, SAFAS_ANA_MPR_MP96_2 22 serialization::archive)

Périmètre thématique des données

- Omics

Droits de propriété intellectuelle

Pôle Métabolome

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

Droits décrits dans les fiches de prestation (<https://nextcloud.inrae.fr/f/35765265>).

Sauf exception:

- Les données et résultats générés sur la plateforme sont la propriété du client/collaborateur/utilisateur/demandeur.
- La plateforme est propriétaire des méthodologies ainsi que de ses développements et adaptations et se réserve le droit d'utiliser une partie des données génériques acquises pour l'annotation de spectres dans le cadre d'autres projets, ou pour des méta-analyses, avec l'autorisation préalable tant que celles-ci ne sont pas dans le domaine public.
- Les bases de données et de connaissances, et les protocoles de la Plateforme sont la propriété stricte de la Plateforme.

Pôle Lipidome

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

Pôle Polyphénols

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

Droits fixés pour chaque projet: accord de consortium, contrats, projet ANR, etc.

Pôle Phénotypage Métabolique à haut débit (plateau HitMe) et Microfluidique

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

Les données appartiennent aux utilisateurs de la plateforme (chercheurs, enseignants-chercheurs, post-doctorants, doctorants, etc). Pour les projets collaboratifs, si les gens n'utilisent pas leur données (projet abandonnées, pas de publication, etc) les utilisateurs signent une autorisation pour que la plateforme puisse réutiliser et exploiter leur données.

Confidentialité

Pôle Métabolome

Identification des jeux de données contenant des données confidentielles

Données de prestations et documentation associée

La prestation ainsi que la documentation associée sont considérées comme des informations confidentielles pour une durée indéterminée (comme indiqué dans la fiche de prestation: <https://nextcloud.inrae.fr/f/35765265>). Fermé le temps du projet, peut être rendu publique à la fin du projet uniquement sur accord et au cas par cas.

Les bases de données et de connaissances, et les protocoles de la Plateforme sont la propriétés stricte de la Plateforme et à ce jours et donc confidentiels.

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

Obligations de confidentialité dès la signature de la fiche de prestation.
Accès au stockage restreints au membre de l'équipe.

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 ?

Pôle Lipidome

Identification des jeux de données contenant des données confidentielles

Données accessible par le personnel et les utilisateurs de la plateforme
Métadonnées des projets (permettant l'exploitation des données) limitées au personnel de la PF ou au porteur de projet.

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

Question sans réponse.

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 ?

Question sans réponse.

Pôle Polyphénols

Identification des jeux de données contenant des données confidentielles

Données accessible par le personnel et les utilisateurs de la plateforme
Métadonnées des projets (permettant l'exploitation des données) limitées au personnel de la PF ou au porteur de projet.

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

Question sans réponse.

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 ?

Question sans réponse.

Pôle Phénotypage Métabolique à haut débit (plateau HitMe) et Microfluidique

Identification des jeux de données contenant des données confidentielles

Les données sont confidentielles jusqu'à publication par l'auteur.

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

Données brutes et finalisées stockées sur le NAS BFP, accès intranet seulement et uniquement par l'unité
Données de traitement accessibles par PLATO par projet mais sans aucune informations/métadonnées

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

Pôle Métabolome

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

Dans le cadre de prestations, les données générées appartiennent au demandeur et, sauf accord, ne sont donc pas diffusables par défaut.

Les données générées dans le cadre d'un projet financé en partie ou en totalité pour un financeur publique suit la politique d'ouverture des données selon le principe "ouvert autant que possible, fermé autant que nécessaire".

Cas particuliers dans le cadre de projets financés par des entreprises privées:

- période d'embargo sur les données

- données jugées "sensibles" pouvant avoir un impact économique et donc soumises à restriction d'accès c.à.d limitées aux seuls partenaires des projets concernés

Quelles sont les réutilisations potentielles de ces données ?

- annotation de spectres
- méta-analyses

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

En règle générale non grâce à l'utilisation de format ouvert (.mzML, .csv, etc).

Cas particuliers :

- * Macro-commandes nécessitant le logiciel NMRProcFlow (open source)
- * Workflow / notebooks nécessitant le logiciel Jupyter (open source)

Comment les données seront-elles partagées ?

Publication sur le portail Data INRAE (<https://data.inrae.fr/>), zenodo (<https://zenodo.org/>), sur des portails dédiés tel que MetaboLights (<https://www.ebi.ac.uk/metabolights/>) ou liés à la publication

Partage sur le portail ODAM (<https://pmb-bordeaux.fr/dataexplorer/>), associé avec un fichier datapackage (format JSON, cf) dans Data INRAE (cf <https://inrae.github.io/ODAM/>)

Partage au sein de MetaboHUB via la plateforme NextCloud INRAE (<https://nextcloud.inrae.fr/>)

Partage de gré à gré (clef USB, email, FileSender, etc) notamment pour la restitution des résultats.

Avec qui ?

- Autre

Données stockées sur le NAS BFP et partagées avec les membres de l'unité BFP

Données partagées publiquement (OpenAccess) quand c'est possible via le portail "Data INRAE" (<https://data.inrae.fr/>), sur des portails dédiés tel que MetaboLights (<https://www.ebi.ac.uk/metabolights/>), ODAM (<https://pmb-bordeaux.fr/dataexplorer/>), ou liés à la publication.

Sous quelle licence ?

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

Pas de licence positionnée automatiquement (indiquée à la publication)

Licence dépendante du contexte et des partenaires (contraintes liés à des partenaires privés, à des consortiums ou des projets).

Pôle Lipidome

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

Partage au sein de l'équipe avant publication.
Publication au gré du partenaire.

Quelles sont les réutilisations potentielles de ces données ?

Question sans réponse.

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

Les données brutes nécessite le logiciel constructeur propriétaire.
Les données peuvent être exportés au format ouvert (mzml, csv, etc).

Comment les données seront-elles partagées ?

Données en cours de traitement sur CNRS MyCore.
Publication sur le portail Data INRAE (<https://data.inrae.fr/>), zenodo (<https://zenodo.org/>) ou sur des portails dédiés.

Avec qui ?

Question sans réponse.

Sous quelle licence ?

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

Pas de licence positionnée automatiquement (indiquée à la publication)

Licence dépendante du contexte et des partenaires (contraintes liés à des partenaires privés, à des consortiums ou des projets).

Pôle Polyphénols

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

Partage au sein de l'équipe avant publication.

Publication au gré du partenaire.

Cas particuliers dans le cadre de projets financés par des entreprises privées:

- période d'embargo sur les données
- données jugées "sensibles" pouvant avoir un impact économique et donc soumises à restriction d'accès c.à.d limitées aux seuls partenaires des projets concernés

Quelles sont les réutilisations potentielles de ces données ?

- Comparaison possible de profils de différents vins afin d'authentifier l'origine
- annotation de spectres
- méta-analyses

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

En règle générale non grâce à l'utilisation de format ouvert (.mzML, .csv, etc).

Comment les données seront-elles partagées ?

Publication sur le portail Data INRAE (<https://data.inrae.fr/>), sur des portails dédiés tel que MetaboLights (<https://www.ebi.ac.uk/metabolights/>) ou lié à la publication

Partage au sein de MetaboHUB via la plateforme NextCloud INRAE (<https://nextcloud.inrae.fr/>)

Partage de gré à gré (clef USB, email, FileSender, etc) notamment pour la restitution des résultats.

Avec qui ?

- Autre

Données stockées sur le NAS de l'équipe oenologie et partagées avec ses membres.

Sous quelle licence ?

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

Pas de licence positionnée automatiquement (indiquée à la publication)

Licence dépendante du contexte et des partenaires (contraintes liés à des partenaires privés, à des consortiums ou des projets).

Pôle Phénotypage Métabolique à haut débit (plateau HitMe) et Microfluidique

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

Restriction d'utilisation tant que les données n'ont pas été publiées.

Quelles sont les réutilisations potentielles de ces données ?

Réutilisation des données potentielle pour de la fouille de données

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

Les données brutes utilise un format propriétaire (.amp96, SAFAS_ANA_MPR_MP96_2 22 serialization::archive) et nécessite le logiciel constructeur (spectrophotomètre).

Les données finales sont directement exploitables (.csv, .xlsx).

Comment les données seront-elles partagées ?

Données stockés sur le NAS BFP (partage au sein de l'unité)

Partage de gré à gré (clef USB, email, FileSender, etc)

Publication sur le portail Data INRAE (<https://data.inrae.fr/>), sur des portails dédiés tel que MetaboLights (<https://www.ebi.ac.uk/metabolights/>) ou lié à la publication

Avec qui ?

- Autre

Données stockées sur le NAS BFP et partagées avec les membres de l'unité BFP

Données partagées publiquement (OpenAccess) quand c'est possible via le portail "Data INRAE" (<https://data.inrae.fr/>), sur des portails dédiés tel que MetaboLights (<https://www.ebi.ac.uk/metabolights/>) ou liées à la publication.

Sous quelle licence ?

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

Organisation et documentation des données

Pôle Métabolome

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

De nombreuses étapes interviennent de l'acquisition des données à la publication, très dépendantes des analyses réalisées (pré-traitement, traitement via Metaboanalyst.ca, workflow4metabolomic, MSDial, NMRProcFlow, BioStatFlow, R, etc).

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 ?

La gestion de métadonnées ainsi que leur format dépendent étroitement du canal de diffusion (en aval), mais à ce jour, il n'y a pas de gestion des métadonnées systématique en amont.

Des métadonnées descriptives (avec ontologies) sont associées lors de la diffusion (partage) des données : ex: Data INRAE, MetaboLights

Des métadonnées structurelles (avec ontologies) sont associées lors de la diffusion (partage) des données via ODAM (<https://inrae.github.io/ODAM/json-schema/>) et liées aux données lors du dépôt dans Data INRAE.

Une partie des métadonnées peut se trouver dans le matériel et méthodes des publications.

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)
métadonnées descriptives	saisie manuelle	Ontologies (POC, CHEBI), Thesaurus (LOV INRAE)	
métadonnées structurelles	saisie manuelle	Ontologies (POC, CHEBI), Thesaurus (LOV INRAE)	

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

Description de l'expérience.

Description des protocoles (expérimentaux, analytiques et bioinformatiques)

Documentation via rendu / compte-rendu de la prestation ou publication

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

Numéro de prestation "MET" = MET-<année>-<P pour prestation><numéro de prestation>-<initiale demandeur>
Exemples: MET-2015-P04-HW, MET-2021-P12-SE (classement automatique par année)

Poste de travail LC-MS

- 1 projet = 1 dossier (répertoire nom = numéro prestation MET)
 - Organisation libre

Poste de travail RMN

- 1 projet = 1 dossier (répertoire nom = numéro prestation MET)
 - Répertoire nmr (1 dossier par spectre: code-numeroechantillon)
 - Répertoire npf (NmrProcFlow: données générées par NMRProcFlow + données zip de données)
 - Répertoire de documents (scientifique, projet, etc)
 - Répertoire Bsf (BioStatFlow: généré par BioStatFlo + données entrées)
 - Données générales du projet: poster, publication, tableau de traitement temporaire, données échantillons, compilation des mesures, table de conversion / fiche de rosette (liaisons expérience - échantillons), table des pesées, table des mises à pH, etc.

Poste d'acquisition MS:

- Nom dossier = <numéro-MET>
 - Pos

- Sequence
- Methode
- Rawdata
- Mzml-data
- Neg
 - Sequence
 - Methode
 - Rawdata
 - Mzml-data

NAS BFP LC-MS

smb://bavo-bfp22.inra.local/vol1/PMB/MET-HD/Archives/Archives LCMS/Données brutes

- Nom dossier: <code numéro MET>
 - Data
 - Methode
 - Séquence (list / ordre d'injection)

NAS EABX:

- dossier qualité-tof
 - Raw data
 - Année
 - projet
- data traité (généré par lles logiciels)
 - Année
 - projet
- Metabofile (metadata + résultat)
 - Année
 - projet

AgroDataRing

sftp://bavo-adr-bfp-01/adr/inra/bap/bfp/archive/

- MET-DJ (dossier des données analysées par Daniel J.)
- MET-LCMS (données LC-MS)
- MET-NMR (données NMR)
- PMB-RedMine (sauvegarde redmine)

Quel est le processus de contrôle qualité des données ?

Question sans réponse.

Pôle Lipidome

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

Données acquises via les différents équipements. Une copie de sauvegarde par jour est réalisé.

Données traitées via les logiciels constructeurs.

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 expériences peuvent être consignées dans une fiche (sous format .xls) résumant l'objectif, le matériel utilisé, les conditions expérimentales, l'emplacement des données brutes et l'analyse des résultats obtenus. Pour l'instant, ces données/métadonnées ne suivent pas d'ontologies ou de vocabulaires contrôlés.

Des métadonnées sont associées lors de la publication (Zenodo, etc).

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

Question sans réponse.

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

Pas de suivi de versions, ni de convention de nommage.

Organisation de l'espace sur MyCore:

- Dossier "Presta"
 - fichier Excel de suivi (status, qui, que quoi, n°MAMA, etc.)
 - Chaque prestation a son dossier (nom = LIP.. = nom de prestation spécifique au pôle)
Numéro LIP = un projet MAMA
Mise à dispo = nom court du projet = un projet MAMA = même nom de dossier sur les appareils utilisés
- Dossier "Mise à dispo"
 - fichiers excel de suivi des projets (qui, quel projet, durée/date, finance)
- Dossier archive classé par année

Quel est le processus de contrôle qualité des données ?

Question sans réponse.

Pôle Polyphénols

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

De nombreuses étapes interviennent de l'acquisition des données à la publication très dépendantes de l'analyse réalisées

- Typique: pré-traitement avec TopSpin et traitement avec Mnova

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 ?

Métadonnées des domaines viticole et oenologique principalement. Pas d'utilisation d'ontologies.

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

Description de l'expérience.
Documentation via rendu / compte-rendu de la prestation ou publication

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

Pas d'organisation particulière.

Quel est le processus de contrôle qualité des données ?

Question sans réponse.

Pôle Phénotypage Métabolique à haut débit (plateau HitMe) et Microfluidique

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

Données générés via les équipements du plateau technique HitMe.

Données brutes stockés sur le NAS BFP (smb://bavo-bfp22.inra.local/vol1/BF/BF\$/HitMelab/).

Données en partie traités via le serveur web LINDA (<https://services.cbib.u-bordeaux.fr/plato/>) suivant le protocole "Measurement of Enzyme Activities and Optimization of Continuous and Discontinuous Assays": <https://doi.org/10.1002/cppb.20003>

Les données sont ensuite exportés de LINDA et stockés sur le NAS BFP

NAS BFP, dossier HitMelab:

- Un dossier par année
 - Un sous-dossier par personne/projet

Au bout de ~2 ans: données archivées (déplacé dans dossier "Archive", i.e. pas supprimées).

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 ?

Question sans réponse.

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

La connaissance/description de l'expérience est nécessaire pour pouvoir analyser et réutiliser les données.

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

Sur le NAS BFP, arborescence:

- Année
 - Un dossier par projet/personne (organisation livre dans le dossier)

Pas de contrôle des versions, ni de conventions de nommage.

Quel est le processus de contrôle qualité des données ?

Les équipements générant les données sont régulièrement contrôlés (campagne de mesure/vérification/étalonnage)

Les données d'optimisations (fichier .xlsx) sont tracés dans le dossier projet (NAS BFP)

Pas de contrôle automatisé sur les données générées

Stockage et sécurité des données

Pôle Métabolome

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

Poste d'acquisition (lié au appareils tel que le spectromètre)

Stocke temporairement les données acquises

Stockage AgroDataRing (<https://datapartage.inrae.fr/Generer/Stocker-les-donnees/AgroDataRing>)

Sauvegarde les données issues du poste d'acquisition (RMN + LC-MS) et données RMN traités.

Stockage NAS BFP (smb://bavo-bfp22.inra.local/vol1/PMB/MET-HD)

Copie des données LC-MS avant suppression du poste d'acquisition

Les données RMN n'y sont pas stockées

Stockage NAS EABX (bxdata)

Données générés par N. Creusot (EABX)

Poste de travail du personnel

Traitement des données et rapports

W4M (<https://workflow4metabolomics.usegalaxy.fr/>)

Traitement des données (stockage temporaire)

Gestion documentaire: RESANA (<https://resana.numerique.gouv.fr/>)

Stockage des procédures (accès restreints)

Publication des données quand c'est possible

Quelle est la volumétrie actuelle et prévisionnelle ?

NAS BFP (PMB / MET-HD, données LC-MS principalement)
Utilisé (depuis 2014, en date de 22/12/2021): 10 To
Volumétrie: ~ 10 Go / projet, voire ~ 100Go pour les gros projets
Prévision: ~ +1To / an
AgroDataRing:
Utilisé (en date de 22/12/2021): 3To
Prévision: < +1To / an
NAS EABX
Utilisé (en 2021): 1,6 To

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...*

Charte des infrastructures de recherche - INRAE: https://www.inrae.fr/sites/default/files/charte_ir.pdf
Pas encore de politique de sécurité des systèmes d'information (PSSI) INRAE.

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 ?

Pas d'échanges avec des acteurs tiers par défaut. Échange de gré à gré au besoin.

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

NAS BFP (<https://nextcloud.inrae.fr/f/51982359>):

- Sauvegarde des données
Serveur principal: snapshots plusieurs fois par jour (toutes les 4h) conservé « aussi longtemps que possible » (un espace est réservé aux snapshots et si cet espace sature, les snapshots les plus anciens sont supprimés).
Copie sur serveur secondaire 2 fois par jour, via Robocopy
Backup sur serveur Linux Media OpenMediaVault (ZFS) 1 fois par semaine
- Surveillance/Alerte en cas de problème (disque, système, onduleur, température, humidité), arrêt automatique en cas de problème électrique
- Mise à jour firmware toutes les semaines (si mise à jour), système en continu (en heures non-ouvrés)
- Antivirus intégré + rapport journalier
- Scanner de vulnérabilité en cours de déploiement pour détecter/corriger les failles de sécurité potentielles

AgroDataRing (<https://datapartage.inrae.fr/Gerer/Stocker-les-donnees/AgroDataRing>)

- Accès restreints par utilisateur/groupe, par SFTP
- Stockage sur disque (ZFS raidz2, équivalent RAID-6, tolérance de panne de 2 disques)
- Instantanés (snapshots) :
Tous les 4h, rétention : 1 semaine
Tous les jours, rétention : 1 mois
Tous les dimanches, rétention : 1 an
- Synchronisation du dossier "archive" sur un site distant, 1 fois par semaine
- Surveillance des disques et des transferts

Pôle Lipidome

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

(Faire éventuellement un lien vers un schéma)

Générations des données sur les instruments:

- Copie tous les jours sur un poste de travail manuellement
- Copie sur disque dur externe tous les mois (~2To)

Donnés du projet stockés sur MyCore@CNRS

Quelle est la volumétrie actuelle et prévisionnelle ?

Données projet de quelques Go

Données brutes: 2To générés (sur ~ 5 ans)

Volumétrie prévisionnelle: ~1To / an

**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...***

UMR CNRS / UB. Les règles et sécurité du système d'information dépends donc de ces deux tutelles.

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 ?

Partage au sein de l'équipe uniquement

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

Pas d'engagements sur la sauvegarde/sécurité des données. Copie de sauvegarde ~1 fois par semaine.

MyCore est sauvegardé (restauration accessible directement) + copie locale sur poste de travail

Pôle Polyphénols

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

Poste d'acquisition (lié au appareils tel que le spectromètre)

Stocque temporairement les données acquises

Sauvegarde les données issues du poste d'acquisition (RMN + LC-MS) et données RMN procédés sur le NAS de l'équipe.

Publication des données quand c'est possible

Quelle est la volumétrie actuelle et prévisionnelle ?

NAS Unité oenologie
Utilisé (depuis 2016, en date de 22/12/2021): 3 To / 12To - Disponible 75%
Prévision: ~ +500Go / an

**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...***

https://www.inrae.fr/sites/default/files/charte_ir.pdf
Pas encore de politique de sécurité des systèmes d'information (PSSI) INRAE.

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 ?

Pas d'échanges avec des acteurs tiers par défaut. Échange de gré à gré au besoin.

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

Pas de mesure de protection pour suivre la production et l'analyse des données

Pôle Phénotypage Métabolique à haut débit (plateau HitMe) et Microfluidique

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

Stockage NAS BFP (smb://bavo-bfp22.inra.local/vol1/BF/BF\$/HitMelab/):
Serveur web PLATO (<https://services.cbib.u-bordeaux.fr/plato/>)

Quelle est la volumétrie actuelle et prévisionnelle ?

Stocké sur le NAS BFP, dossier HitMelab (de 2010-2021) (22/12/2021): ~ 40 Go

- Volumétrie: ~50 Mo par projet
- Prévisionnelle: ~ +1Go /an

**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...***

Charte des infrastructures de recherche - INRAE: https://www.inrae.fr/sites/default/files/charte_ir.pdf

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 ?

Pas d'échanges avec des acteurs tiers. Échange de gré à gré au besoin.

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

Données stockées sur le **NAS BFP** (<https://nextcloud.inrae.fr/f/51982359>):

- Sauvegarde des données
Serveur principal: snapshots plusieurs fois par jour (toutes les 4h) conservé « aussi longtemps que possible » (un espace est réservé aux snapshots et si cet espace sature, les snapshots les plus anciens sont supprimés).
Copie sur serveur secondaire 2 fois par jour, via Robocopy
Backup sur serveur Linux Media OpenMediaVault (ZFS) 1 fois par semaine
- Surveillance/Alerte en cas de problème (disque, système, onduleur, température, humidité), arrêt automatique en cas de problème électrique
- Mise à jour firmware toutes les semaines (si mise à jour), système en continu (en heures non-ouvrés)
- Antivirus intégré + rapport journalier
- Scanner de vulnérabilité en cours de déploiement pour détecter/corriger les failles de sécurité potentielles

Données de traitement stockés sur le serveur web **PLATO**

Offre hébergement: <https://nextcloud.inrae.fr/f/55730952> (lien accès restreints)

Une sauvegarde journalière est réalisée, conservée 2 semaines.

Une sauvegarde hebdomadairement le dimanche, conservée 12 semaines.

Archivage et conservation des données

Pôle Métabolome

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

A moyen terme, les données des prestations sont conservées (raw data, peak list, metadata, etc).

Il peut y avoir jusqu'à 5 à 7 ans entre l'acquisition et la publication.

Il n'y a pas de suppression.

Conservation des données traitées:

- LC-MS: sur le NAS BFP (dossier personnel ou de la PF), à la fin du projet ou à saturation des espaces de travail.
- RMN: sur poste de travail et AgroDataRing

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 ?

Pas d'archivage pérenne. Simple conservation sur le NAS BFP ou EABX.

Utilisation des plateformes nationales pour les données publiées (data.inrae, MetaboLight, etc).

Quelle est la durée de conservation des données ?

La durée minimale de conservation des données est la durée de la prestation.
Pas de durée de rétention maximale définie.

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

Fonctionnement de l'unité BFP (NAS BFP) et financement MetaboHUB.

Pôle Lipidome

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

Les données publiées sont conservées (en s'appuyant sur la structure: zenodo, data.inrae.fr, etc).
Il n'y a pas de politique de conservation en dehors des données publiées.

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 ?

Plateforme Zenodo principalement.

Quelle est la durée de conservation des données ?

Question sans réponse.

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

Question sans réponse.

Pôle Polyphénols

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

A moyen terme, les données des prestations sont conservées (raw data, peak list, metadata, etc).
Il peut y avoir jusqu'à 5 à 7 ans entre l'acquisition et la publication.
Il n'y a pas de suppression.

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 ?

Pas d'archivage pérenne. Simple conservation sur le NAS de l'équipe

Quelle est la durée de conservation des données ?

La durée minimale de conservation des données est la durée de la prestation.
Pas de durée de rétention maximale définie.

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

Fonctionnement de l'unité œnologie (NAS) et financement MetaboHUB.

Pôle Phénotypage Métabolique à haut débit (plateau HitMe) et Microfluidique

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

Pas de suppression. Données de faible volumétrie permettant la conservation sans grand besoin de stockage.

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 ?

Pas d'archivage pérenne. Simple conservation sur le NAS BFP

Quelle est la durée de conservation des données ?

Pas de suppression pour l'instant.

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

Fonctionnement de l'unité BFP (NAS BFP)

