
Data Management Plan - Metabolic Profiling and Metabolomic Platform P2M2

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

Renseignements sur le plan

Titre du plan	Data Management Plan - Metabolic Profiling and Metabolomic Platform P2M2
Version	04/05/2024
Objet/périmètre du plan	<p>This document aims to ensure the quality, traceability, and reusability of the data collected and generated by the Metabolic Profiling and Metabolomic Platform (P2M2). The specific objectives of this DMP include:</p> <ol style="list-style-type: none">1. Data collection and storage: defines the data collection and storage modalities for metabolomics data, including file formats, metadata to collect, data backup and protection strategies.2. Data management: specifies the procedures for managing metabolomics data, including documentation of experimental protocols, unique identifiers for each sample and analysis, and data access and sharing rights management.3. Data analysis and interpretation: describes the methods for analyzing and interpreting metabolomics data, including the software used, processing algorithms, data normalization, and quality.4. Data archiving and sharing: defines the modalities for archiving and sharing metabolomics data, including existing data repositories or national or international archives, data sharing policies, and data usage licenses.5. Training and awareness: provides for training and awareness-raising activities for metabolomics platform users on good practices for managing metabolomics data, as well as on policies and regulations related to data use and sharing.
Domaines de recherche (selon classification de l'OCDE)	Agricultural biotechnology, Agriculture, forestry, and fisheries, Biological sciences (Natural sciences), Chemical engineering, Chemical sciences, Computer and information sciences
Langue	eng
Date de création	2023-05-03
Date de dernière modification	2023-05-04
Identifiant	https://doi.org/10.57745/JBPH2Q
Type d'identifiant	DOI
Licence	Etalab Open License 2.0

Renseignements sur le projet

Titre du projet Data Management Plan - Metabolic Profiling and Metabolomic Platform P2M2

Acronyme P2M2

Résumé The P2M2 (Metabolic Profiling and Metabolomic Platform) is a state-of-the-art analytical facility dedicated to the comprehensive and quantitative analysis of small molecules in complex biological systems, with a focus on plants for agronomic traits of interest. The platform employs cutting-edge mass spectrometry-based metabolomics technologies and bioinformatics tools to provide a wide range of metabolomics services to academic and industrial users studying plants. The P2M2 platform offers high-throughput metabolomic profiling services for various plant tissues, including leaves, roots, seeds, and fruits, with the goal of identifying and quantifying metabolites that play a crucial role in plant growth, development, and stress response. The platform is equipped with a range of mass spectrometers and chromatography systems that allow for the analysis of a broad range of plant metabolites. In addition, the platform integrates a software development activity in bioinformatics to better structure its data and aid in the understanding of acquisition data. This helps to improve the management and processing of data generated by the platform, ensuring high-quality and reliable results for its users.

Date de début 2022-01-03

Partenaires

- Institut Francais des Productions Cidricoles (<http://www.ifpc.eu/>)
- Institut national de recherche pour l'agriculture, l'alimentation et l'environnement ()
- Université Rennes ()

Produits de recherche :

1. Chromatogramme (RAW Data) (Jeu de données)
2. Analysis report for the client
3. Sample for metabolomics analysis (Objet physique)
4. Method/Application applied to a group of samples (Texte)
5. Software development (Logiciel)

Contributeurs

Nom	Affiliation	Rôles
Berardocco Solenne - https://orcid.org/0000-0002-4584-0965	Université de Rennes	<ul style="list-style-type: none"> • Personne contact pour les données (Analysis report)
Bouchereau Alain	Université de Rennes	<ul style="list-style-type: none"> • Coordinateur du projet
Charton Sophie - https://orcid.org/0000-0002-9594-4025	INRAE	<ul style="list-style-type: none"> • Personne contact pour les données (Sample)
Filangi Olivier - https://orcid.org/0000-0002-2094-3271	Université de Rennes	<ul style="list-style-type: none"> • Personne contact pour les données (Software) • Responsable du plan de gestion de données
Guichard Hugues - https://orcid.org/0000-0002-4461-6865	IFPC	
Guyot Sylvain - https://orcid.org/0000-0002-9367-3196	INRAE	
Marnet Nathalie	INRAE	<ul style="list-style-type: none"> • Personne contact pour les données (Chromatogram)
Sotin Hélène	INRAE	<ul style="list-style-type: none"> • Personne contact pour les données (Method)

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.

Data Management Plan - Metabolic Profiling and Metabolomic Platform P2M2

Informations sur la structure

Nom de la structure

The Metabolic Profiling and Metabolomic Platform (P2M2)

Type de structure

- Plateforme, plateau technique
-

Identifiant de la structure

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

<https://doi.org/10.17180/WQV7-X981>

Responsabilités dans la structure

Nom, Prénom	Courriel	Rôle
Alain Bouchereau Sylvain Guyot Rémi Bauduin	alain.bouchereau@universite-rennes.fr sylvaing.guyot@inrae.fr remi.bauduin@ifpc.eu	Management team / Scientific managers
Solenne Berardocco	solenne.berardocco@universite-rennes.fr	Technical Manager and Quality Correspondent
Solenne Berardocco Younès Dello	solenne.berardocco@universite-rennes.fr younes.dello@inrae.fr	Quality managers
Olivier Filangi	olivier.filangi@inrae.fr	IT activity manager

Etablissement(s) tutelle(s)

INRAE, Université de Rennes, L'institut Agro Rennes-Angers, IFPC

Département de rattachement INRAE (ou anciens départements Inra)

- TRANSFORM : Aliments, produits biosourcés et déchets

- BAP : Biologie et amélioration des plantes

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

INRAE

Informations sur le plan de gestion

DOI (*version publiée du plan de gestion*)

<https://doi.org/10.57745/JBPH2Q>

Historique des versions

Date	n° de version	Status	Auteur	Affiliation de l'auteur (se reporter à l' annuaire INRAE)	Validé par	Validé le
04/05/2023	1	beta	Olivier Filangi	IGEPP		

Présentation générale des données

Chromatogramme (RAW Data)

Mode d'obtention des données

- Données générées par la structure

Raw data refers to the output generated by analytical instruments and is typically encoded in a manufacturer-specific format. At P2M2, a range of formats are utilized for the acquisition data generated by its mass spectrometry-based metabolomics technologies, including Xcalibur, GCMS Solution, MassLynx, Open Lab, Empower, Chromeleon, and ChemStation. These acquisition data files contain critical information about the composition of a sample and are initially saved on the local disks of the control PCs. They represent the first step in the analysis pipeline of the P2M2 platform and are used to generate more advanced data processing and analysis.

Origine

- Analyse
-

Type de données

- Dataset

Nature des données

Metabolomics acquisition data refers to the raw data generated by mass spectrometry-based technologies that capture signals representing the small molecules present in a given biological sample. The format of this acquisition data is typically proprietary, encoded in binary format by the instrument manufacturer.

Format des données

The proprietary formats of the acquisition software include Xcalibur, GCMS Solution, MassLynx, Open Lab, Empower, Chromeleon, and ChemStation

Périmètre thématique des données

- Plant Health and Pathology
- Plant Breeding and Plant Products
- Omics
- Food and food processing
- Chemistry and chemical engineering

Analysis report for the client

Mode d'obtention des données

- Données générées par la structure

Handwritten targeted analysis reports are created by statistically processing acquisition data and comparing it with analysis standards.

Origine

- Observation
- Expérimentation

Type de données

- Text

Nature des données

Targeted analysis reports

Format des données

docx

Périmètre thématique des données

- Plant Health and Pathology
- Plant Breeding and Plant Products
- Food and food processing
- Chemistry and chemical engineering

Sample for metabolomics analysis

Mode d'obtention des données

- Données produites par un tiers

Samples are provided by the applicant.

Origine

- Expérimentation

Type de données

- Physical object

Nature des données

different tissues and/or structures such as leaves, stems, roots, flowers, fruits, or seeds.

Format des données

Question sans réponse.

Périmètre thématique des données

- Plant Health and Pathology
- Plant Breeding and Plant Products

Method/Application applied to a group of samples

Mode d'obtention des données

- Données générées par la structure

Acquisition methods to set up GC, LC and MS analysis.

Origine

- Analyse

Type de données

- Software

Nature des données

File

Format des données

Proprietary format

Périmètre thématique des données

- Chemistry and chemical engineering

Software development

Mode d'obtention des données

- Données générées par la structure

The software sources of P2M2 are developed using R, Scala, and TypeScript programming languages.

Origine

- Code
-

Type de données

- Software
-

Nature des données

software sources / text files

Format des données

R, Scala

Périmètre thématique des données

- Computer science

Droits de propriété intellectuelle

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

The ownership of data and information resulting from the processing of a sample belongs to the applicant. However, the P2M2 platform retains institutional control over the software source codes developed by its agents.

Sensibilité des données

Chromatogramme (RAW Data)

Identification du niveau de sensibilité des jeux de données

- Diffusion limitée

The processed data, reports and samples are not published by the platform and their access is limited to the platform members and the applicant.

Quelles sont les mesures prises et les normes auxquelles il est nécessaire de se conformer pour garantir la sécurité des données sensibles ?

The data is accessible uniquely on the platform's internal network or at the client's request via a file transfer system

S'il y a des données à caractère personnel, quelles sont les mesures envisagées pour les protéger au cours du projet ou dans le cadre d'une réutilisation ?

Question sans réponse.

Analysis report for the client

Identification du niveau de sensibilité des jeux de données

- Diffusion limitée

The processed data, reports and samples are not published by the platform and their access is limited to the platform members and the applicant.

Quelles sont les mesures prises et les normes auxquelles il est nécessaire de se conformer pour garantir la sécurité des données sensibles ?

The data is accessible uniquely on the platform's internal network or at the client's request via a file transfer system

S'il y a des données à caractère personnel, quelles sont les mesures envisagées pour les protéger au cours du projet ou dans le cadre d'une réutilisation ?

Question sans réponse.

Sample for metabolomics analysis

Identification du niveau de sensibilité des jeux de données

- Diffusion limitée

The processed data, reports and samples are not published by the platform and their access is limited to the platform members and the applicant.

Quelles sont les mesures prises et les normes auxquelles il est nécessaire de se conformer pour garantir la sécurité des données sensibles ?

The data is accessible uniquely on the platform's internal network or at the client's request via a file transfer system

S'il y a des données à caractère personnel, quelles sont les mesures envisagées pour les protéger au cours du projet ou dans le cadre d'une réutilisation ?

Question sans réponse.

Method/Application applied to a group of samples

Identification du niveau de sensibilité des jeux de données

- Diffusion limitée

The methods are accessible only to members of the platform.

Quelles sont les mesures prises et les normes auxquelles il est nécessaire de se conformer pour garantir la sécurité des données sensibles ?

The data is accessible uniquely on the platform's internal network

S'il y a des données à caractère personnel, quelles sont les mesures envisagées pour les protéger au cours du projet ou dans le cadre d'une réutilisation ?

Question sans réponse.

Software development

Identification du niveau de sensibilité des jeux de données

- Public

The source codes of the platform are published on the GITHUB repository (<https://github.com/p2m2>)

Quelles sont les mesures prises et les normes auxquelles il est nécessaire de se conformer pour garantir la sécurité des données sensibles ?

Question sans réponse.

S'il y a des données à caractère personnel, quelles sont les mesures envisagées pour les protéger au cours du projet ou dans le cadre d'une réutilisation ?

Question sans réponse.

Partage des données

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

There is no requirement to share the data.

Software sources developed in the framework of P2M2 activities are available on GITHUB from the p2m2 organization

Quelles sont les réutilisations potentielles de ces données ?

no reuse of samples, processed data and methods. The software sources are under MIT license in order to be reusable by the academic and industrial community

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

Yes, the acquisition software is required to read the acquisition data and method files: Xcalibur, GCMS Solution, MassLynx, Open Lab, Empower, Chromeleon, and ChemStation
Microsoft Office Suite is used for analysis reports

Comment les données seront-elles partagées ?

Publication on the portal [Data INRAE](#)
Publication On GITHUB

Avec qui ?

- Tous (open acces)
-

Sous quelle licence ?

- Licence ouverte <https://www.etalab.gouv.fr/licence-ouverte-open-licence> (compatible CC-BY)
-

Organisation et documentation des données

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

Utiliser éventuellement un lien vers un schéma illustrant les processus

Metabolomics data acquisition and processing follows a standardized workflow. Raw data are first generated using liquid chromatography (LC) or gas chromatography (GC) and mass spectrometry (MS) techniques. Proprietary acquisition software such as Xcalibur, GCMS Solution, MassLynx, Open Lab, Empower, Chromeleon and ChemStation are used to generate raw data files from the acquisition methods.

Then, the manufacturer's software is used to preprocess the raw data to remove noise and improve data quality.

As part of the targeted analysis, basic statistical analyses are performed using a spreadsheet program to facilitate the formalization of the analysis report. These analysis reports are generated using office software, such as Microsoft Office.

As part of the targeted analysis, the data is subjected to specific workflow treatments (W4M[1]) previously defined with the client.

All these data are initially transferred to a server located in the P2M2 office. An automatic process is used to synchronize this data on a server located in the INRAE Ile de France data center dedicated to the P2M2 activity and accessible to all P2M2 platform agents, and

on the storage server of the IGEPP unit of the Genouest regional bioinformatics platform (host unit) in Rennes to be archived in the INRAE AgroDataRing[2] project.

The software sources are created using open source editors and kept under version control on the GITHUB platform. The development process is organized into different branches, and production versions are clearly identified using tags. The entire software life cycle is managed using the GITHUB software forge.

[1] Franck Giacomoni, Gildas Le Corguillé, Mishari Monsoor, Marion Landi, Pierre Pericard, et al.. Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. *Bioinformatics*, 2015, 31 (9), pp.1493-1495. (10.1093/bioinformatics/btu813). (hal-01123263)

[2] Pierre Adenot, Stéphane Bansard, David Benaben, Veronique Brunaud, Christophe Caron, et al.. AgroDataRing: Une infrastructure partagée et mutualisée pour le stockage longue durée. *Cahier des Techniques de l'INRA*, 2018, pp.1-7. (hal-02097192)

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 ?

Question sans réponse.

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

Data files are organized in projects and are not versioned.
Software sources are organized in projects and are versioned

Quel est le processus de contrôle qualité des données ?

The software sources are subjected to a continuous integration process with the definition of unit test

Stockage et sécurité des données

Les systèmes d'information de la structure ont-ils fait l'objet d'une analyse de risques ou d'une homologation ?

- Non

Quels types de supports physiques sont utilisés pour stocker les données ?

Digital data (raw, reports, methods) are accessible by agents via Nextcloud/Genouest-AgroDataRing and the P2M2 server in the ille de France datacenter.

The software sources are available on GITHUB

Quelles sont les mesures de sécurité mises en place lors des étapes de transfert des données ?

Data transfer uses the SSH (Secure Socket Shell) protocol

Quelle est la volumétrie actuelle et prévisionnelle ?

The storage servers contain 1.5 TB of acquisition data.

L'entité hébergeant physiquement les données a-t-elle une politique de sécurité de l'information et a-t-elle un plan d'assurance sécurité ?

Question sans réponse.

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 ?

Question sans réponse.

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

Question sans réponse.

Les agents de la structure ont-ils bénéficié d'une sensibilisation aux bonnes pratiques d'hygiène numérique ?

- Non
-

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 ?

One needs to have a data archiving policy for acquisition data. However, analysis reports and method files are not as crucial and do not require special archiving measures.

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 ?

AgroDataRing and data.gouv.fr are used for

Quelle est la durée de conservation des données ?

Targeted analyses are subject to a two-year retention period
Non-targeted analyses are retained for the duration of the project (Genouest policy)
Archived data benefit from the lifetime of the AgroDataRing project

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

Question sans réponse.