



**HAL**  
open science

## Plan de Management des Données (DMP) du projet ANR MucoReaDy "Mucoviscidose et dysfonctionnements respiratoires"

Philippe Poncet, Stephanie Delage Santacreu

### ► To cite this version:

Philippe Poncet, Stephanie Delage Santacreu. Plan de Management des Données (DMP) du projet ANR MucoReaDy "Mucoviscidose et dysfonctionnements respiratoires". 2022. hal-03726293

**HAL Id: hal-03726293**

**<https://hal.science/hal-03726293v1>**

Submitted on 27 Feb 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# DMP du projet "Mucoviscidose et dysfonctionnements respiratoires"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - Modèle de PGD (français)" fourni par Agence nationale de la recherche (ANR).

## Renseignements sur le plan

<b>Titre du plan</b>	DMP du projet "Mucoviscidose et dysfonctionnements respiratoires"
<b>Version</b>	Version initiale
<b>Objet/périmètre du plan</b>	Ce Plan de Gestion des Données (PGD ou DMP) est un document vivant faisant l'inventaire de l'ensemble des données produites ou utilisées par le projet ANR MucoReaDy. Il va des données expérimentales (brutes ou post-traitées) et des codes sources de calcul aux documents de communication (publication et site web) en passant par les contrats de sous-traitances.
<b>Domaines de recherche (selon classification de l'OCDE)</b>	Clinical medicine, Mathematics, Computer and information sciences, Health sciences
<b>Langue</b>	fra
<b>Date de création</b>	2022-05-03
<b>Date de dernière modification</b>	2022-07-18
<b>Identifiant</b>	hal-03726293
<b>Type d'identifiant</b>	identifiant local
<b>Licence</b>	Creative Commons Attribution Non Commercial No Derivatives 4.0 International
<b>Documents (publications, rapports, brevets, plan expérimental....), sites web associés</b>	<ul style="list-style-type: none"><li>• Certification du niveau de sécurité du Cloud UPPA pour entreposer les données à partage maîtrisé : <a href="https://www.ssllabs.com/ssltest/analyze.html?d=nuage.univ%2dpau.fr&amp;hideResults=on&amp;latest">https://www.ssllabs.com/ssltest/analyze.html?d=nuage.univ%2dpau.fr&amp;hideResults=on&amp;latest</a></li><li>• Site web du projet MuCoReaDy : <a href="https://sites.google.com/view/mucoready">https://sites.google.com/view/mucoready</a></li></ul>

## Renseignements sur le projet

<b>Titre du projet</b>	Mucoviscidose et dysfonctionnements respiratoires
<b>Acronyme</b>	MucoReaDy
<b>Résumé</b>	L'objectif du projet MucoReaDy est d'étudier l'altération du mécanisme de clairance mucociliaire pour la pathologie de la mucoviscidose et de l'aspergillose bronchopulmonaire allergique (ABPA), dans une perspective de dynamique du mucus. Dans ces pathologies, le dysfonctionnement pulmonaire s'aggrave progressivement au fil du temps, entraînant une diminution du pourcentage du volume expiratoire forcé prévu en 1s (ppFEV1), annonçant une maladie pulmonaire s'aggravant au fil du temps, souvent irréversible. Le dépistage des nouveau-nés a permis de mettre en place rapidement des thérapies standardisées chez les enfants atteints de mucoviscidose, ce qui a permis d'améliorer leur condition de croissance, de nutrition et de maladie

pulmonaire. Pourtant, la mortalité prématurée demeure, avec une survie médiane prévue de 47 ans. Nous souhaitons obtenir une compréhension précise des mécanismes impliqués dans la production de mucus, la maturation, l'auto-stratification et surtout le mouvement induit par ses interactions avec son environnement complexe et mobile. Afin d'explorer ce sujet, nous souhaitons bénéficier d'une synergie entre les mathématiques et la simulation (LMAP, UMR CNRS 5142), la médecine (CHU Bordeaux et CHU Toulouse) tant pour la médecine adulte que pour la pédiatrie, la physique et la chimie de la matière molle (IPREM, UMR CNRS 5254). Les questions que nous abordons pour étudier ce problème sont, parmi beaucoup d'autres : Comment les différents états de maturation du mucus influencent sa rhéologie? En fonction de son environnement, est-il possible de prédire la fonctionnalité du mucus par modélisation (modèle mathématique et simulation). Est-il possible de construire une cartographie prédictive qui ne prendra en compte que quelques paramètres rhéologiques afin de montrer l'évolution de la clairance mucociliaire? En particulier dans le contexte de la pathologie de la mucoviscidose et de son traitement? La méthodologie peut-elle être suffisamment robuste pour estimer le risque d'une résistance à la thérapie médicamenteuse ou non ? Compte tenu des modèles de polymérisation, pouvons-nous expliquer les temps caractéristiques de l'évolution de l'aspergillose ? Pouvons-nous expliquer sa résistance à la thérapie antifongique ? Cela clarifierait le protocole entre la thérapie corticoïde/antifongique pour l'ABPA. Le projet est organisé en 4 tâches principales, la première étant consacrée à la coordination. La deuxième tâche se concentrera sur la modélisation avancée et la simulation numérique de l'environnement de l'épithélium pulmonaire, en impliquant le moins possible de paramètres : système CFTR/transcystose, cils élastiques, cellules gobelets, mucines, maturation du mucus, pus/activité bactérienne, voies aériennes. Au niveau clinique, le projet MucoReaDy veut prouver que ce diagnostic rapide et bon marché peut être utilisé pour vérifier l'efficacité d'une thérapie sur un patient, en quelques semaines seulement au lieu de 3 à 6 mois, échelle de temps clinique actuelle. Ces aspects concernent la tâche 3. En effet, notre but est de vérifier si l'évolution de la rhéologie correspond à l'évolution des indicateurs cliniques du patient, afin d'identifier le plus rapidement possible toute résistance à la thérapie. L'objectif urgent est le correcteur de nouvelle génération pour la FK, la triple thérapie (Trikafta) Elexacaftor-tezacaftor-ivacaftor, disponible en 2020 pour les personnes de plus de 12 ans. Une autre originalité du projet est la modélisation de la prolifération des agents et de leur mobilité dans cet environnement, traitée dans la tâche 4. L'objectif est d'obtenir les facteurs de croissance, de transport et de diffusion/dispersion l'échelle cellulaire, en vue de les injecter dans des modèles à plus grande échelle. Cela sera appliqué à la prolifération fongique et bactérienne, en se concentrant sur l'ABPA mais aussi appliqué à *Pseudomonas Aeruginosa* et au complexe *Burkholderia cepacia*.

### Sources de financement

- Agence Nationale de la Recherche : ANR-20-CE45-0022

**Date de début** 2021-02-01

**Date de fin** 2025-01-31

## Partenaires

- Institut des sciences analytiques et de physico-chimie pour l'environnement et les matériaux (200711919E)
- CHU Bordeaux-Pellegrin, Pôle Pédiatrie (201822917L)
- CHU Toulouse-Larrey, Direction de la Recherche (199221775H)
- Laboratoire de Mathématiques et de leurs Applications de Pau - UMR CNRS 5142 (200511822H)

## Produits de recherche :

1. Code prototype et librairie de calcul (Logiciel)
2. Données massives issues du calcul intensif (Jeu de données)
3. Tableaux cliniques CHU Toulouse et Bordeaux (Jeu de données)
4. Données rhéologiques brutes et post-traitées (Jeu de données)
5. Informations légales: notices aux patients et documents contractuels (Texte)
6. Publications: ACM et autres diffusions scientifiques (Texte)

## Contributeurs

Nom	Affiliation	Rôles
DELAGE SANTACREU Stéphanie		<ul style="list-style-type: none"><li>• Responsable du plan de gestion de données</li></ul>
PONCET Philippe	Laboratoire de mathématiques et de leurs applications - Pau	<ul style="list-style-type: none"><li>• Coordinateur du projet</li><li>• Personne contact pour les données (Données Rhéo, Code Prototype, Publications, Informations légales, Données calcul, Excel CHU)</li></ul>

# DMP du projet "Mucoviscidose et dysfonctionnements respiratoires"

## 1. Description des données et collecte ou réutilisation de données existantes

### Code prototype et librairie de calcul

#### 1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Les données de cette étude sont produites par un code de calcul original à partir de la plateforme HySoP. Cette plateforme de calcul hybride GPU-CPU HySoP est développée par LMAP-LJK-CNAM, est open-source et sera utilisée prioritairement.

#### 1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

La plateforme HySoP est multi-langage : Python, C/C++, Fortran 77 et 90, OpenCL.

\* Les tableaux cliniques seront saisis en format tableur (excel protégé par mot de passe).

\* Le rhéomètre produira des données sous format texte brut (fréquence d'oscillation du cône vs vitesse résultante) et sera immédiatement converti en caractéristiques rhéologiques (modules  $G'$  et  $G''$ , viscosité induite, cisaillement, temps de relaxation, seuil, ...), faisant quelques Ko.

\* Le code de calcul produira des données destinées à être post-traitées. Les données brutes au format binaire vont de l'ordre de 10Mo (10 sorties de 10Mo environ) pour le développement et test, jusqu'à 1To (100 sorties de 10Go) en production à échelle fine.

\* La rédaction collaborative (articles et comptes rendus) sera effectuée sur le ShareLaTeX du CNRS (INSMI).

\* La plateforme HySoP est multi-langage : Python, C/C++, Fortran 77 et 90, OpenCL.

### Données massives issues du calcul intensif

#### 1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Les données produites par le code de calcul HySoP là sont originales et ne sont pas basés sur des données pré-existantes.

#### 1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Le code de calcul produira des données destinées à être post-traitées. Les données brutes au format binaire vont de l'ordre de 10Mo (10 sorties de 10Mo environ) pour le développement et test, jusqu'à 1To (100 sorties de 10Go) en production à échelle fine.

### Tableaux cliniques CHU Toulouse et Bordeaux

#### 1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Les tableaux cliniques de cette étude seront issus de saisies manuelles: ces points-là sont originaux et ne sont pas basés sur des données pré-existantes.

**1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?**

Les tableaux cliniques seront saisis en format tableur (excel protégé par mot de passe).

**Données rhéologiques brutes et post-traitées**

**1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?**

Les données expérimentales produites par un rhéomètre sont originales et ne sont pas basés sur des données pré-existantes.

**1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?**

Le rhéomètre produira des données sous format texte brut (fréquence d'oscillation du cône vs vitesse résultante) et sera immédiatement converti en caractéristiques rhéologiques (modules G' et G'', viscosité induite, cisaillement, temps de relaxation, seuil, ...), faisant quelques Ko.

**Informations légales: notices aux patients et documents contractuels**

**1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?**

Les documents d'information auprès des patients et leur signature de non opposition à l'étude sont basées à partir d'études précédentes menées au CHU Toulouse et adaptées à la pédiatrie pour le CHU Bordeaux.

**1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?**

Ces données sont sur support imprimé.

**Publications: ACM et autres diffusions scientifiques**

**1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?**

Question sans réponse.

### 1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

La rédaction collaborative (articles et comptes rendus) sera effectuée sur le ShareLaTeX du CNRS (INSMI).

## 2. Documentation et qualité des données

### 2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

Chaque partenaire collectera ses propres données et sera responsable de l'organisation et de l'indexation de ses données.

### 2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

- Pour les données rhéologiques, il y aura alicotage des échantillons (lorsque c'est possible) afin de procéder à des mesures multiples,
- Le code de calcul HySoP est soumis à un contrôle de version et une gestion stricte des branches de développement afin de garantir un fonctionnement sans dégradation irréversible.

## 3. Stockage et sauvegarde pendant le processus de recherche

### 3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

- Les données médicales pseudo-anonymisées (tableaux cliniques) et données rhéologiques seront déposées sur le cloud sécurisé nuage.univ-pau.fr
- Les blocs de données issues du calcul intensif, de tailles modérées et non pertinentes à partagées, seront hébergées sur le cluster de calcul pyrene.univ-pau.fr sur un stockage avec redondance. Le grand volume de stockage nécessaire est un service financé pour 5 ans par le projet au titre de l'utilisation de ressources de calcul et stockage mutualisé.
- L'archivage (stockage froid) sera assuré par le projet iRods de Nouvelle-Aquitaine avec réplication des données à Bordeaux, Pau et Poitiers.

### 3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

Les données à partager entre les membres du projet, quel que soit le type de donnée produite, seront exclusivement stockées sur le cloud Nuage de l'uppa (nuage.univ-pau.fr) qui est évalué Sécurité A+ par le service Qualys de SSL Labs:

<https://www.ssllabs.com/ssltest/analyze.html?d=nuage.univ%2dpau.fr&hideResults=on&latest>

Les données médicales pseudo-anonymisées (tableaux cliniques) y seront déposées au format excel protégé par mot de passe et ayant une sauvegarde incrémentale à haute disponibilité.

Les données produites par le rhéomètre seront également déposées dans cet environnement sécurisé (leur production initiale sur le rhéomètre sera effacée car produite dans un environnement non sécurisé, mais n'embarquant pas de données personnelles).

Les blocs de données issues du calcul intensif, de tailles modérées et non pertinentes à partagées, seront hébergées sur le cluster de calcul pyrene.univ-pau.fr dans un environnement Linux gérant les droits de manière stricte. Le transfert vers le stockage froid sera assuré par protocole SSH.

## 4. Exigences légales et éthiques, codes de conduite

### Code prototype et librairie de calcul

4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Pas de données personnelles pour le développement, déploiement et utilisation en production du code de calcul.

4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Le code de calcul HySoP est initialement prévu en licence GPL 2.0, mais ce choix va évoluer avant mise en ligne de la version partagée. Cette discussion aura lieu en concertation entre l'UPPA (LMAP Pau), l'UGA (LJK Grenoble) et le CNAM (Paris), co-développeurs de HySoP.

4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Tous les développements du coeur de HySoP sont réalisés en concertation avec les co-développeurs, les branches spécifiques pour les projets de recherche étant décrites au sein du serveur Git hébergeant la totalité de HySoP et de ses branches.

### Données massives issues du calcul intensif

4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Sans objet.

4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Les données sont propriétés intellectuelle de la personne les ayant produit (par déontologie et non par contractualisation) jusqu'à leur publication où elle deviennent publiques.

4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Afin de faire respecter la propriété intellectuelle mentionnée ci-dessus, la responsabilité appartient à la personne ayant produit ces données de protéger leur accès.



## Tableaux cliniques CHU Toulouse et Bordeaux

**4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?**

Les tableaux cliniques contiennent des données personnelles semi-anonymisées. Les médecins et ARC (Attaché de Recherche Clinique) en charge de ces patients seront les seuls à même de pouvoir remonter aux noms des patients (RGPD satisfaites). Le fichier semi-anonymisé sera protégé par mot de passe et sur un cloud sécurisé.

**4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?**

Les fichiers sont propriété des CHU les ayant produits, et mis à disposition de l'ensemble des partenaires du projet.

**4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?**

Chaque patient dont les données sont présentes dans ce fichier aura signé un document de non opposition et aura eu un document d'information (conformément à la législation, loi Jardé, cf. section Informations légales).

## Données rhéologiques brutes et post-traitées

**4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?**

Les données rhéologiques seront obtenues par deux moyens, à partir des échantillons de mucus fournis par les deux CHU:  
- Par mesure directe à l'IPREM, auquel cas les données sont exprimées par la référence anonymisée du patient (conforme aux RGPD),  
- Par le prestataire Rhéonova, dont la gestion des échantillons est contractualisée et a été validée par les juristes des deux CHU comme conforme aux RGPD.

**4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?**

Les données résultantes de la rhéométrie sont en partagées avec les partenaires du projet et seront publiées.

**4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?**

Les patients ayant produit le mucus mesuré par rhéométrie ont été informé de la démarche et ont signé leur non-opposition.

## Informations légales: notices aux patients et documents contractuels

**4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à**

#### caractère personnel et sur la sécurité des données sera-t-il assuré ?

Les patients des CHU produisant du mucus destiné à être analysé (rhéométrie et bactériologie) ont été informé du but du projet et ont signé une non opposition. Ces deux documents ont été validés par les juristes des CHU Toulouse et Bordeaux, et leur conformité avec les RGPD et la loi Jardé a été établi.

#### 4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Les documents soumis au patient (information et non-opposition) sont propriété des CHU les ayant produits, bien que leur forme et contenu soient partagés entre les partenaires du projet.

#### 4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Les juristes des CHU Toulouse-Larrey et Bordeaux-Pellegrin ont établi que le Comité de protection des personnes (CPP, anciennement comité d'éthique) n'avait pas à être saisi et que le protocole établi suffit pour la conformité avec la loi Jardé.

### Publications: ACM et autres diffusions scientifiques

#### 4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Les données publiées seront totalement anonymisées, dans l'esprit de notre publication "Numerical and experimental investigation of mucociliary clearance breakdown in cystic fibrosis. *Journal of Biomechanics*, Elsevier, 2017, 53, pp.56-63".

#### 4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Question sans réponse.

#### 4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

La question déontologique la plus épineuse pour les publications et de définir qui seront les auteurs et dans quel ordre. A cette fin, les premiers résultats de mucoready étant attendu en octobre 2022, les quatre partenaires du projet et les juristes des CHU s'entendront fin septembre 2022 et rédigeront un document informel précisant tous ces éléments. Ce document sera moins strict qu'un accord formel de consortium, puisque tous les partenaires sont des institutions publiques.

## 5. Partage des données et conservation à long terme

**5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?**

Les données concernant des tableaux cliniques ou mesures individualisées ne seront jamais partagées (RGPD).

Certains résultats des codes de calcul seront rendus totalement disponibles dans des entrepôts de données publics, sauf ceux pouvant conduire à un dépôt de brevet.

---

**5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?**

L'ensemble des données (et méta-données associées) seront conservées en stockage froid redondant sur le iRodS régional Nouvelle-Aquitaine (triple redondance Pau-Bordeaux-Poitiers).

---

**5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?**

Les tableaux de données "patients" seront au format Excel protégé par mot de passe (RGPD) en plus d'être soumis à un accès restreint et protégé (nuage.univ-pau.fr).

---

**5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?**

Cette partie sera à préciser pendant l'exécution du projet.

---

## **6. Responsabilités et ressources en matière de gestion des données**

**6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?**

Pour les données médicales, les responsables scientifique de MucoReaDy des CHU Bordeaux (Dr Stéphanie Bui) et Toulouse (Dr Marlène Murriss) seront responsables des données produites par leur services. La manipulation et l'édition de ces fichiers sera assurée par 2 ARC (*Attaché de Recherche Clinique*, un par CHU) dédiés au projet.

Les données rhéologiques seront gérées par Anthony Laffore, Ingénieur à l'IPREM et personnes réalisant les mesures. Jean-Matthieu Etancelin (MCF LMAP-UPPA) sera en charge des données de production du code de calcul HySoP.

---

**6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?**

Les données destinées à être partagées seront organisées de manière FAIR lors des publications associées à la méthodologie ou à l'exploitation de ces données.