

---

## DMP du projet "Gestionnaire de données adaptatif multi-niveau pour l'Exascale"

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

### Renseignements sur le plan

<b>Titre du plan</b>	DMP du projet "Gestionnaire de données adaptatif multi-niveau pour l'Exascale"	
<b>Domaines de recherche (selon classification de l'OCDE)</b>	Health biotechnology, Industrial biotechnology	
<b>Langue</b>	fra	
<b>Date de création</b>	2022-12-07	
<b>Date de dernière modification</b>	2022-12-08	
<b>Identifiant</b>	12222	
<b>Type d'identifiant</b>	DOI	
<b>Licence</b>	<b>Nom</b>	Creative Commons Zero v1.0 Universal
	<b>URL</b>	<a href="http://spdx.org/licenses/CC0-1.0.json">http://spdx.org/licenses/CC0-1.0.json</a>

### Renseignements sur le projet

**Titre du projet** Gestionnaire de données adaptatif multi-niveau pour l'Exascale

**Acronyme** ADMIRE

**Résumé** La nécessité de traiter des ensembles de données de plus en plus volumineux est l'une des principale raison motivant la construction de systèmes HPC Exascale. Cependant, les architectures de stockage uniformes que l'on retrouve dans les systèmes HPC classiques ne répondent plus aux exigences de performances des applications de traitement intensif des données. L'accès aux fichiers non coordonné associé à une bande passante limitée font du système de fichiers parallèle centralisé un facteur limitant des performances. Dans le même temps, on assiste à l'émergence de hiérarchies de stockage multi-niveaux qui offre la possibilité de surmonter ce point de blocage. Mais l'optimisation des performances vient au prix d'un contrôle fin de la congestion et d'équilibrage de charge entre le calcul et les performances de stockage. Malheureusement, il n'existe pas interfaces conçues de manière appropriées pour gérer la pile d'E / S dans ce contexte de gestion des ressources. L'objectif principal du projet ADMIRE est de réaliser ce contrôle fin en créant une pile d'E / S arbitrant dynamiquement entre les besoins de calcul et ceux de stockage grâce à une orchestration globale, la malléabilité du calcul et des E / S, et la gestion des ressources de stockage à tous les niveaux de la hiérarchie de stockage. Pour y parvenir, nous développerons environnement défini par logiciel basé sur les principes de contrôle commande évolutifs, cet environnement sera conçu pour passer à l'échelle avec des canaux pour le contrôle et les données séparés, ainsi que par une orchestration des composants clés du système et des applications via des points de contrôle pré-définis. Notre solution purement logicielle permettra d'augmenter considérablement le débit des systèmes HPC et les performances des applications individuelles - et par conséquent de réduire la consommation d'énergie - en tirant parti de niveaux de stockage de nœuds locaux rapides et a consommation énergétique optimisées, d'un système de fichiers Européen ad-hoc et de capacité de traitement en transit / in situ. De plus, notre pile augmentée d'E / S offrira de la qualité de service (QoS) et de la résilience. Un prototype intégré et opérationnel sera validé avec plusieurs cas d'utilisation de divers domaines, notamment le climat / météo, les sciences de la vie, la physique, la télédétection et l'apprentissage profond.

**Sources de financement**

- Agence Nationale de la Recherche : ANR-20-EHPC-0003
- European Commission : 123456
- Institut National de la Santé et de la Recherche Médicale : 56987

**Date de début** 2022-12-15

**Partenaires**

- Centre de Recherche Inria Bordeaux - Sud-Ouest
- PARATOOLS SAS

**Produits de recherche :**

1. jeu de données 1 (Jeu de données)
2. New research output 2
3. New research output 3

## Contributeurs

Nom	Affiliation	Rôles
Busin Anne - <a href="https://orcid.fr">https://orcid.fr</a>	INIST - 198822	
cfjr gmail		<ul style="list-style-type: none"> <li>• Personne contact pour les données (jd1)</li> </ul>
Cosserat Françoise - <a href="https://orcid.fr">https://orcid.fr</a>	INIST - 123456	<ul style="list-style-type: none"> <li>• Coordinateur du projet</li> <li>• Personne contact pour les données (Research Output 2, Research Output 3)</li> <li>• Responsable du plan de gestion de données</li> </ul>
free fco		<ul style="list-style-type: none"> <li>• Responsable du stockage des données (jd1)</li> </ul>

## Budget

Type de coût (Titre)	Montant	Etape du cycle de vie
Frais de personnel (Cout 1)	120 EUR	Coûts liés à la collecte/production des données - <b>jd1</b>
Stockage (Estimation du coût de stockage et calcul auprès du mésocentre MESO@LR)	55.0 EUR	Coût total de stockage et/ou calcul (coût calculé automatiquement) - <b>jd1</b>

Droits d'auteur :

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

# DMP du projet "Gestionnaire de données adaptatif multi-niveau pour l'Exascale"

---

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

### jeu de données 1

#### 1.1 Description générale du produit de recherche

<b>Nom</b>	jeu de données 1
<b>Type</b>	Jeu de données
<b>Mots clés</b>	<ul style="list-style-type: none"><li>• protection de la foret (Agrovoc)</li></ul>
<b>Mots clés (texte libre)</b>	dsfq
<b>Langue</b>	fra
<b>Date de publication</b>	2022-12-21
<b>Contient des données personnelles ?</b>	Oui
<b>Contient des données sensibles ?</b>	Non
<b>Prend en compte des aspects éthiques ?</b>	Ne sais pas

#### 1.2 Est-ce que des données existantes seront réutilisées ?

Question sans réponse.

#### 1.3 Comment seront produites/collectées les nouvelles données ?

<b>Coûts</b>	<ul style="list-style-type: none"><li>• Frais de personnel : 120 EUR</li></ul>
--------------	--

### New research output 2

#### 1.1 Description générale du produit de recherche

<b>Nom</b>	New research output 2
------------	-----------------------

**1.2 Est-ce que des données existantes seront réutilisées ?**

Question sans réponse.

**1.3 Comment seront produites/collectées les nouvelles données ?**

Question sans réponse.

**New research output 3**

**1.1 Description générale du produit de recherche**

**Nom** New research output 3

**1.2 Est-ce que des données existantes seront réutilisées ?**

Question sans réponse.

**1.3 Comment seront produites/collectées les nouvelles données ?**

Question sans réponse.

**2. Documentation et qualité des données**

**jeu de données 1**

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

**Standards de métadonnées/données**

- AVM (Astronomy Visualization Metadata) : <https://rdamsc.bath.ac.uk/msc/m3>

---

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

---

**New research output 2**

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

Question sans réponse.

---

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

Question sans réponse.

---

**New research output 3**

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

Question sans réponse.

---

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

Question sans réponse.

**3. Exigences légales et éthiques, code de conduite**

## **jeu de données 1**

**3.2 Quelles sont les contraintes juridiques (sensibilité des données autres qu'à caractère personnel, confidentialité, ...) à prendre en compte pour le partage et le stockage des données ?**

Question sans réponse.

---

**3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?**

Question sans réponse.

## **New research output 2**

**3.2 Quelles sont les contraintes juridiques (sensibilité des données autres qu'à caractère personnel, confidentialité, ...) à prendre en compte pour le partage et le stockage des données ?**

Question sans réponse.

---

**3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?**

Question sans réponse.

## **New research output 3**

**3.2 Quelles sont les contraintes juridiques (sensibilité des données autres qu'à caractère personnel, confidentialité, ...) à prendre en compte pour le partage et le stockage des données ?**

Question sans réponse.

---

### 3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?

Question sans réponse.

## 4. Traitement et analyse des données

### jeu de données 1

#### 4.1 Comment et avec quels moyens seront traitées les données ?

Question sans réponse.

### New research output 2

#### 4.1 Comment et avec quels moyens seront traitées les données ?

Question sans réponse.

### New research output 3

#### 4.1 Comment et avec quels moyens seront traitées les données ?

Question sans réponse.

## 5. Stockage et sauvegarde des données pendant le processus de recherche

### jeu de données 1



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

### Description

blabla

### Identification de la ressource

- MESO@LR : <https://cat.opidor.fr/index.php/MESO@LR>

### Demande de ressources de stockage et calcul à MESO@LR (en To)

**Type de service** Stockage inférieur à 100 To

**Justification** blabla justificatif

**Taille des données** 120

**Unité** To

**Durée de stockage en mois** 12

**Unité** mois

### Politique de sauvegarde

- Politique de sauvegarde de MESO@LR : Votre espace est sauvegardé par snapshots et réplication. Les snapshots sont des copies des données à des moments réguliers et prédéfinis, permettant de revenir en arrière en cas de fausse manipulation sur un fichier. Les snapshots ne vont stocker que les différences sur les fichiers modifiés depuis le dernier snapshot. La réplication permet une sauvegarde complète des données. Cette copie des données se fait dans une salle physiquement distante du stockage initial. Elle permet une récupération des données lorsque toutes les données sont détruites

### Mesures de sécurité

L'offre de service propose la réplication des données, cette copie conforme se situe dans une salle distance du stockage primaire. Complémentaire à cette réplication, la possibilité d'avoir des snapshots selon la politique choisie offre une excellente protection contre les ransomwares. La connexion se fait selon votre solution d'authentification souhaitée : Active Directory ou LDAP, sinon par le LDAP de Meso. Le Mésocentre n'est pas encore habilité à recevoir des données de santé.

### Coût total de stockage et/ou calcul (coût calculé automatiquement)

- Stockage : 55,0 EUR

## New research output 2

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

Question sans réponse.

## New research output 3

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

Question sans réponse.

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

### jeu de données 1

#### 6.1 Comment les données seront-elles partagées ?

##### Entrepôt/Catalogue de données

- CDS : <https://cat.opidor.fr/index.php/CDS> (CoreTrustSeal/WDS/Centre de référence thématique Astronomie et astrophysique de Recherche Data Gouv)

##### Caractéristiques des fichiers déposés

- fichier 1 (text/csv)

#### 6.2 Comment les données seront-elles conservées à long terme ?

Question sans réponse.

### New research output 2

#### 6.1 Comment les données seront-elles partagées ?

Question sans réponse.

#### 6.2 Comment les données seront-elles conservées à long terme ?

Question sans réponse.

### New research output 3

#### 6.1 Comment les données seront-elles partagées ?

Question sans réponse.

---

**6.2 Comment les données seront-elles conservées à long terme ?**

Question sans réponse.