



**HAL**  
open science

## Plan de gestion de données du projet PaléoLASER

Thibault Clauzel

► **To cite this version:**

Thibault Clauzel. Plan de gestion de données du projet PaléoLASER. Centre national de la recherche scientifique. 2024. hal-04684839

**HAL Id: hal-04684839**

**<https://hal.science/hal-04684839>**

Submitted on 3 Sep 2024

**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 "PALÉOLASER"

*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

Titre du plan	DMP du projet "PaléoLASER"
Version	Version finale
Objet/périmètre du plan	Ce plan vise à définir et identifier les modes de production des données afin de mieux préparer leur diffusion.
Domaines de recherche (selon classification de l'OCDE)	Earth and related environmental sciences
Langue	fra
Date de création	2023-02-07
Date de dernière modification	2024-09-02

## RENSEIGNEMENTS SUR LE PROJET

Titre du projet	PaléoLASER
Acronyme	Application de la VCOF-CRDS à l'étude des environnements du passé
Résumé	L'anomalie d'oxygène-17 ( $\Delta 17O$ ) dans les minéraux carbonatés est un traceur émergent, susceptible d'offrir des informations précieuses sur les paléo-environnements, les mécanismes de bio-calciification et l'histoire géologique de la Terre. La mesure précise de $\Delta 17O$ dans ces carbonates par spectrométrie de masse pose de sérieux problèmes analytiques, mais les progrès récents de méthodes spectroscopiques comme la VCOF-CRDS (« V-shaped Cavity Optical Feedback - Cavity Ring-Down Spectroscopy ») permettent aujourd'hui de contourner ces problèmes. C'est pourquoi la Plateforme Analytique Géosciences Paris-Saclay (PANOPLY) se dote en 2023 d'un instrument VCOF-CRDS unique au monde, capable de mesurer $\Delta 17O$ dans le CO <sub>2</sub> et les carbonates avec une précision <10 ppm, une simplicité et une rapidité inégalées. Le succès de cette entreprise repose sur le transfert d'un savoir-faire technique des physiciens aux géochimistes et sur l'adaptation des procédures existantes pour tenir compte des exigences de l'analyse géochimique. Le but du projet est d'accompagner l'installation de la VCOF-CRDS à PANOPLY et de réaliser les premières applications géochimiques et paléoenvironnementales, qui porteront sur la caractérisation du fractionnement $17O/16O$ à l'équilibre entre eau et calcite, sur l'étude des déséquilibres isotopiques dans les bio-carbonates, sur les reconstructions paléo-climatiques à partir de spéléothèmes, et sur l'évolution de la composition isotopique des océans à l'échelle du Phanérozoïque.

Sources de financement

- Centre National de la Recherche Scientifique :

Date de début 2023-01-01

Date de fin 2024-06-30

Partenaires

- Laboratoire des Sciences du Climat et de l'Environnement UMR 8212 (200611689J)
- Institut des sciences de la Terre Paris (200918450V)
- Géosciences Paris-Saclay (200412804E)
- Laboratoire Interdisciplinaire de Physique (201122569N)

Produits de recherche :

1. Default research output

Contributeurs

Nom	Affiliation	Rôles
Clauzel Thibault		<ul style="list-style-type: none"> <li>• Personne contact pour les données</li> <li>• Responsable du plan de gestion de données</li> </ul>
Daerön Mathieu <a href="https://orcid.org/0000-0003-1210-9786">https://orcid.org/0000-0003-1210-9786</a>	Laboratoire des Sciences du Climat et de l'Environnement	<ul style="list-style-type: none"> <li>• Coordinateur du projet</li> </ul>

## DMP DU PROJET "PALÉOLASER"

### 1. DESCRIPTION DES DONNÉES ET COLLECTE OU RÉUTILISATION DE DONNÉES EXISTANTES

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

Les nouvelles données brutes acquises durant ce projet seront des données de compositions isotopiques acquises par des logiciels de spectrométrie de masse. Les données brutes sont acquises sous un format universel de style .txt et seront ensuite traitées et corrigées sur des logiciels de traitement de données (Excel, R, Python) puis enregistrées sous format .csv ou .xls

Des données préexistantes dans des publications scientifique pourront être exploitées si nécessaire, afin de discuter des résultats obtenus par la méthode de VCOF-CRDS.

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

Les données numériques (en .txt, .csv ou .xls) seront collectées et produites durant l'ensemble du projet. Le volume de données estimé est d'environ 1000 à 2000 par an.

Les codes des programmes utilisés pour les traitements de données sont mis en ligne et accessibles à tous sur Zenedo.

### 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 ?

Les données seront accompagnées des métadonnées suivantes :

- date d'acquisition
- utilisateur

- méthodologie utilisée
  - données liées aux protocoles préalables de traitement chimique.
- L'ensemble de ces métadonnées seront associées dans un même fichier unique permettant d'avoir une traçabilité fine de la production des données sur les différents échantillons.

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

De nombreux contrôles de qualité des données sont effectués :

- Calibration des données
- Estimation du bruit de chaque mesure
- Répétabilité du signal
- Reproductibilité des mesures
- Mesures de nombreux standards internationaux
- Publication des données et revue par les pairs

### **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 ?

- Synchronisation des données et sauvegarde automatique sur plusieurs serveurs dont certains en lecture seule.
- Sauvegarde sur ordinateur personnel et disque dur externe.
- Partage des données entre les différents participants du projet via le système de partage Syncthing.
- Les données brutes, codes de traitement et données traitées seront publiées sur Zenodo lors de la publication des résultats.

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

Pas de données sensibles sur ce projet.

### **4. EXIGENCES LÉGALES ET ÉTHIQUES, CODES DE CONDUITE**

*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 à caractère personnel traitées dans le cadre de ce projet.

*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 ?*

Pour le post-doctorat, financé à 100% par le CNRS, la législation qui s'applique est celle des agents CNRS.

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

Pas de questions éthiques ou déontologiques sur ce projet.

## **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 seront partagées après le projet, sans restrictions, via des publications dans des articles scientifiques qui seront déposés sur des plateformes publiques type HAL. Celles-ci seront donc disponibles après publication, pour des raisons de compétitivité vis-à-vis de l'élaboration de la méthode.

Les données pourront être utilisables par tous.

Les programmes et codes de traitement utilisés seront également partagés sur Zenodo.

*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 acquises qui démontrent le fonctionnement de l'outil mis en place seront publiées sur Zenodo lors de la publication.

Les publications scientifiques associées au projet seront archivées dans HAL.

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

Les fichiers au format .csv et .xls pourront être utilisés à l'aide des logiciels de traitement de données classiques (Excel, Open office) et exploités à l'aide d'outils de traitement variés.

Les codes produits sont utilisables via Python.

*5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PÉRENNE (COMME LE DOI) SERA-T-ELLE ASSURÉE POUR CHAQUE JEU DE DONNÉES ?*

Les données publiées dans les articles bénéficieront d'un DOI attribués par les éditeurs de revues.

Lorsque les publications seront déposées dans HAL, chacune se verra également attribuer un identifiant unique.

## **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) ?*

Dr Mathieu Daerön, Laboratoire des Sciences du Climat et de l'Environnement (UMR 8212 mixte CEA/CNRS/UVSQ)

*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) ?*

L'objectif même de ce projet est de s'assurer de la validité des données obtenues par spectrométrie laser et de rendre la méthode accessible aux autres utilisateurs. À ce titre, une partie temporelle d'environ 5% du travail est dédiée à s'assurer que les données sont bien FAIR (une majeure partie du processus est automatisé). Le tri et la sélection des données accessibles mobilise l'ensemble des collaborateurs du projet PaléoLASER. Les données essentielles dans le cadre du projet seront publiées dans des articles scientifiques, ce qui en assurera leur disponibilité, tandis que les autres données seront rendues accessibles sur Zenodo.